

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>C12N 15/52, 15/82, 5/10, 1/21, C12P 7/64, C11C 1/00, C07K 14/405, 14/28, A01H 5/00</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 00/42195</b> <b>(43) International Publication Date:</b> 20 July 2000 (20.07.00)
<b>(21) International Application Number:</b> PCT/US00/00956 <b>(22) International Filing Date:</b> 14 January 2000 (14.01.00)  <b>(30) Priority Data:</b> 09/231,899 14 January 1999 (14.01.99) US  <b>(71) Applicant:</b> CALGENE, LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).  <b>(72) Inventors:</b> FACCIOITTI, Daniel; 2636 Lafayette Drive, Davis, CA 95616 (US). METZ, James, George; 2830 Belhaven Place, Davis, CA 95616 (US). LASSNER, Michael; 721 Falcon Avenue, Davis, CA 95616 (US).  <b>(74) Agent:</b> RAE-VENTER, Barbara; Rae-Venter Law Group, P.C., P.O. Box 60039, Palo Alto, CA 94306 (US).		<b>(81) Designated States:</b> BR, CA, IL, JP, MX, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> SCHIZOCHYTRIUM PKS GENES  <b>(57) Abstract</b>  The present invention relates to compositions and methods for preparing poly-unsaturated long chain fatty acids in plants, plant parts and plant cells, such as leaves, roots, fruits and seeds. Nucleic acid sequences and constructs encoding PKS-like genes required for the poly-unsaturated long chain fatty acid production, including the genes responsible for eicosapentenoic acid production of <i>Shewanella putrefaciens</i> and novel genes associated with the production of docosahexenoic acid in <i>Vibrio marinus</i> are used to generate transgenic plants, plant parts and cells which contain and express one or more transgenes encoding one or more of the PKS-like genes associated with such long chain poly-unsaturated fatty acid production. Expression of the PKS-like genes in the plant system permits the large scale production of poly-unsaturated long chain fatty acids such as eicosapentenoic acid and docosahexenoic acid for modification of the fatty acid profile of plants, plant parts and tissues. Manipulation of the fatty acid profiles allows for the production of commercial quantities of novel plant oils and products.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## SCHIZOCHYTRIUM PKS GENES

5

### INTRODUCTION

#### 10 Field of the Invention

This invention relates to modulating levels of enzymes and/or enzyme components capable of modifying long chain poly-unsaturated fatty acids (PUFAs) in a host cell, and constructs and methods for producing PUFAs in a host cell. The invention is exemplified by production of eicosapentenoic acid (EPA) using genes derived from *Shewanella putrefaciens* and *Vibrio marinus*.

#### Background

Two main families of poly-unsaturated fatty acids (PUFAs) are the  $\omega$ 3 fatty acids, exemplified by eicosapentenoic acid, and the  $\omega$ 6 fatty acids, exemplified by arachidonic acid. PUFAs are important components of the plasma membrane of the cell, where they can be found in such forms as phospholipids, and also can be found in triglycerides. PUFAs also serve as precursors to other molecules of importance in human beings and animals, including the prostacyclins, leukotrienes and prostaglandins. Long chain PUFAs of importance include docosahexenoic acid (DHA) and eicosapentenoic acid (EPA), which are found primarily in different types of fish oil, gamma-linolenic acid (GLA), which is found in the seeds of a number of plants, including evening primrose (*Oenothera biennis*), borage (*Borago officinalis*) and black currants (*Ribes nigrum*), stearidonic acid (SDA), which is found in marine oils and plant seeds, and arachidonic acid (ARA), which along with GLA is found in filamentous fungi. ARA can be purified from animal tissues including liver and adrenal gland. Several genera of marine bacteria are known which synthesize either EPA or DHA. DHA is present in human milk along with ARA.

PUFAs are necessary for proper development, particularly in the developing infant brain, and for tissue formation and repair. As an example, DHA, is an important constituent of many human cell membranes, in particular nervous cells (gray matter), muscle cells, and spermatozoa and believed to affect the development of brain functions in general and to be essential for the development of eyesight. EPA and DHA have a number of nutritional and pharmacological uses. As an example adults affected by diabetes (especially non insulin-dependent) show



deficiencies and imbalances in their levels of DHA which are believed to contribute to later coronary conditions. Therefore a diet balanced in DHA may be beneficial to diabetics.

For DHA, a number of sources exist for commercial production including a variety of marine organisms, oils obtained from cold water marine fish, and egg yolk fractions. The purification of DHA from fish sources is relatively expensive due to technical difficulties, making DHA expensive and in short supply. In algae such as *Amphidinium* and *Schizochytrium* and marine fungi such as *Thraustochytrium* DHA may represent up to 48% of the fatty acid content of the cell. A few bacteria also are reported to produce DHA. These are generally deep sea bacteria such as *Vibrio marinus*. For ARA, microorganisms including the genera *Mortierella*, *Entomophthora*, *Phytium* and *Porphyridium* can be used for commercial production. Commercial sources of SDA include the genera *Trichodesma* and *Echium*. Commercial sources of GLA include evening primrose, black currants and borage. However, there are several disadvantages associated with commercial production of PUFAs from natural sources. Natural sources of PUFA, such as animals and plants, tend to have highly heterogeneous oil compositions. The oils obtained from these sources can require extensive purification to separate out one or more desired PUFA or to produce an oil which is enriched in one or more desired PUFA.

Natural sources also are subject to uncontrollable fluctuations in availability. Fish stocks may undergo natural variation or may be depleted by overfishing. Animal oils, and particularly fish oils, can accumulate environmental pollutants. Weather and disease can cause fluctuation in yields from both fish and plant sources. Cropland available for production of alternate oil-producing crops is subject to competition from the steady expansion of human populations and the associated increased need for food production on the remaining arable land. Crops which do produce PUFAs, such as borage, have not been adapted to commercial growth and may not perform well in monoculture. Growth of such crops is thus not economically competitive where more profitable and better established crops can be grown. Large-scale fermentation of organisms such as *Shewanella* also is expensive. Natural animal tissues contain low amounts of ARA and are difficult to process. Microorganisms such as *Porphyridium* and *Shewanella* are difficult to cultivate on a commercial scale.

Dietary supplements and pharmaceutical formulations containing PUFAs can retain the disadvantages of the PUFA source. Supplements such as fish oil capsules can contain low levels of the particular desired component and thus require large dosages. High dosages result in ingestion of high levels of undesired components, including contaminants. Care must be taken in providing fatty acid supplements, as overaddition may result in suppression of endogenous biosynthetic pathways and lead to competition with other necessary fatty acids in various lipid fractions *in vivo*, leading to undesirable results. For example, Eskimos having a diet high in  $\omega$ 3 fatty acids have an increased tendency to bleed (U.S. Pat. No. 4,874,603). Fish oils have

unpleasant tastes and odors, which may be impossible to economically separate from the desired product, such as a food supplements. Unpleasant tastes and odors of the supplements can make such regimens involving the supplement undesirable and may inhibit compliance by the patient.

A number of enzymes have been identified as being involved in PUFA biosynthesis. Linoleic acid (LA, 18:2  $\Delta$  9, 12) is produced from oleic acid (18:1  $\Delta$  9) by a  $\Delta$ 12-desaturase. GLA (18:3  $\Delta$  6, 9, 12) is produced from linoleic acid (LA, 18:2  $\Delta$  9, 12) by a  $\Delta$ 6-desaturase. ARA (20:4  $\Delta$  5, 8, 11, 14) is produced from DGLA (20:3  $\Delta$  8, 11, 14), catalyzed by a  $\Delta$ 5-desaturase. Eicosapentenoic acid (EPA) is a 20 carbon, omega 3 fatty acid containing 5 double bonds ( $\Delta$  5, 8, 11, 14, 17), all in the *cis* configuration. EPA, and the related DHA ( $\Delta$  4, 7, 10, 13, 16, 19, C22:6) are produced from oleic acid by a series of elongation and desaturation reactions. Additionally, an elongase (or elongases) is required to extend the 18 carbon PUFAs out to 20 and 22 carbon chain lengths. However, animals cannot convert oleic acid (18:1  $\Delta$  9) into linoleic acid (18:2  $\Delta$  9, 12). Likewise,  $\mu$ -linolenic acid (ALA, 18:3  $\Delta$  9, 12, 15) cannot be synthesized by mammals. Other eukaryotes, including fungi and plants, have enzymes which desaturate at positions  $\Delta$ 12 and  $\Delta$ 15. The major poly-unsaturated fatty acids of animals therefore are either derived from diet and/or from desaturation and elongation of linoleic acid (18:2  $\Delta$  9, 12) or  $\mu$ -linolenic acid (18:3  $\Delta$  9, 12, 15).

Poly-unsaturated fatty acids are considered to be useful for nutritional, pharmaceutical, industrial, and other purposes. An expansive supply of poly-unsaturated fatty acids from natural sources and from chemical synthesis are not sufficient for commercial needs. Because a number of separate desaturase and elongase enzymes are required for fatty acid synthesis from linoleic acid (LA, 18:2  $\Delta$  9, 12), common in most plant species, to the more saturated and longer chain PUFAs, engineering plant host cells for the expression of EPA and DHA may require expression of five or six separate enzyme activities to achieve expression, at least for EPA and DHA, and for production of quantities of such PUFAs additional engineering efforts may be required, for instance the down regulation of enzymes competing for substrate, engineering of higher enzyme activities such as by mutagenesis or targeting of enzymes to plastid organelles. Therefore it is of interest to obtain genetic material involved in PUFA biosynthesis from species that naturally produce these fatty acids and to express the isolated material alone or in combination in a heterologous system which can be manipulated to allow production of commercial quantities of PUFAs.

#### Relevant Literature

Several genera of marine bacteria have been identified which synthesize either EPA or DHA (DeLong and Yayanos, *Applied and Environmental Microbiology* (1986) 51: 730-737). Researchers of the Sagami Chemical Research Institute have reported EPA production in *E. coli* which have been transformed with a gene cluster from the marine bacterium, *Shewanella*

*putrefaciens*. A minimum of 5 open reading frames (ORFs) are required for fatty acid synthesis of EPA in *E. coli*. To date, extensive characterization of the functions of the proteins encoded by these genes has not been reported (Yazawa (1996) *Lipids* 31, S-297; WO 93/23545; WO 96/21735).

5       The protein sequence of open reading frame (ORF) 3 as published by Yazawa, USPN 5,683,898 is not a functional protein. Yazawa defines the protein as initiating at the methionine codon at nucleotides 9016-9014 of the *Shewanella* PKS-like cluster (Genbank accession U73935) and ending at the stop codon at nucleotides 8185-8183 of the *Shewanella* PKS-like cluster. However, when this ORF is expressed under control of a heterologous promoter in an *E.*  
10 *coli* strain containing the entire PKS-like cluster except ORF 3, the recombinant cells do not produce EPA.

Polyketides are secondary metabolites the synthesis of which involves a set of enzymatic reactions analogous to those of fatty acid synthesis (see reviews: Hopwood and Sherman, *Annu. Rev. Genet.* (1990) 24: 37-66, and Katz and Donadio, in *Annual Review of Microbiology* (1993)  
15 47: 875-912). It has been proposed to use polyketide synthases to produce novel antibiotics (Hutchinson and Fujii, *Annual Review of Microbiology* (1995) 49:201-238).

### **SUMMARY OF THE INVENTION**

Novel compositions and methods are provided for preparation of long chain poly-  
20 unsaturated fatty acids (PUFAs) using polyketide-like synthesis (PKS-like) genes in plants and plant cells. In contrast to the known and proposed methods for production of PUFAs by means of fatty acid synthesis genes, by the invention constructs and methods are provided for producing PUFAs by utilizing genes of a PKS-like system. The methods involve growing a host cell of interest transformed with an expression cassette functional in the host cell, the expression  
25 cassette comprising a transcriptional and translational initiation regulatory region, joined in reading frame 5' to a DNA sequence to a gene or component of a PKS-like system capable of modulating the production of PUFAs (PKS-like gene). An alteration in the PUFA profile of host cells is achieved by expression following introduction of a complete PKS-like system responsible for a PUFA biosynthesis into host cells. The invention finds use for example in the  
30 large scale production of DHA and EPA and for modification of the fatty acid profile of host cells and edible plant tissues and/or plant parts.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

Figure 1 provides designations for the ORFs of the EPA gene cluster of *Shewanella*.  
35 Figure 1A shows the organization of the genes; those ORFs essential for EPA production in *E. coli* are numbered. Figure 1B shows the designations given to subclones.

Figure 2 provides the *Shewanella* PKS-like domain structure, motifs and 'Blast' matches of ORF 6 (Figure 2A), ORF 7 (Figure 2B), ORF 8 (Figure 2C), ORF 9 (Figure 2D) and ORF 3 (Figure 2E). Figure 2F shows the structure of the region of the Anabeana chromosome that is related to domains present in *Shewanella* EPA ORFs.

5 Figure 3 shows results for pantethenylation - ORF 3 in *E. coli* strain SJ16. The image shows [ $C^{14}$ ]  $\beta$ -Alanine labelled proteins from *E. coli* (strain SJ16) cells transformed with the listed plasmids. Lane 1 represents pUC19, lane 2 represents pPA-NEB ( $\Delta$  ORF 3), lane 3 represents pAA-Neb (EPA+), lane 4 represents ORF 6 subclone, lane 5 represents ORF 6 + ORF 3 subclones, and lane 6 represents ORF 3 subclone. ACP and an unknown (but previously  
10 observed) 35 kD protein were labelled in all of the samples. The high molecular mass proteins detected in lanes 2 and 5 are full-length (largest band) and truncated products of the *Shewanella* ORF-6 gene (confirmed by Western analysis). *E. Coli* strain SJ16 is conditionally blocked in  $\beta$ -alanine synthesis.

Figure 4A shows the DNA sequence (SEQ ID NO:1) for the PKS-like cluster found in  
15 *Shewanella*, containing ORF's 3-9. Figure 4B shows the amino acid sequence (SEQ ID NO:2) of ORF 2, which is coded by nucleotides 6121-8103 of the sequence shown in Fig 4A. Figure 4C shows the amino acid sequence (SEQ ID NO:3) of the published, inactive ORF3, translated from the strand complementary to that shown in Figure 4A, nucleotides 9016-8186. Figure 4D shows the nucleotide sequence 8186-9157 (SEQ ID NO:4); its complementary strand codes for  
20 ORF 3 active in EPA synthesis. Figures 4E-J show the amino acid sequences (SEQ ID NOS:5-10) corresponding to ORF's 4-9, which are encoded by nucleotides 9681-12590 (SEQ ID NO:81), 13040-13903 (SEQ ID NO:82), 13906-22173 (SEQ ID NO:83), 22203-24515 (SEQ ID NO:84), 24518-30529 (SEQ ID NO:85) and 30730-32358 (SEQ ID NO:86), respectively, of Figure 4A. Figure 4K shows the amino acid sequence (SEQ ID NO:11) corresponding to  
25 nucleotides 32834-34327.

Figure 5 shows the sequence (SEQ ID NO:12) for the PKS-like cluster in an approximately 40 kb DNA fragment of *Vibrio marinus*, containing ORFs 6, 7, 8 and 9. The start and last codons for each ORF are as follows: ORF 6: 17394, 25352; ORF 7: 25509, 28160; ORF 8: 28209, 34265; ORF 9: 34454, 36118.

30 Figure 6 shows the sequence (SEQ ID NO:13) for an approximately 19 kb portion of the PKS-like cluster of Figure 5 which contains the ORFs 6, 7, 8 and 9. The start and last codons for each ORF are as follows: ORF 6: 411, 8369 (SEQ ID NO:77); ORF 7: 8526, 11177 (SEQ ID NO:78); ORF 8: 11226, 17282 (SEQ ID NO:79); ORF 9: 17471, 19135 (SEQ ID NO:80).

35 Figure 7 shows a comparison of the PKS-like gene clusters of *Shewanella putrefaciens* and *Vibrio marinus*; Figure 7B is the *Vibrio marinus* operon sequence.

Figure 8 is an expanded view of the PKS-like gene cluster portion of *Vibrio marinus* shown in Figure 7B showing that ORFs 6, 7 and 8 are in reading frame 2, while ORF 9 is in reading frame 3.

Figure 9 demonstrates sequence homology of ORF 6 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 6 is depicted on the vertical axis, and the *Vibrio* ORF 6 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity. The repeated lines in the middle correspond to the multiple ACP domains found in ORF 6.

Figure 10 demonstrates sequence homology of ORF 7 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 7 is depicted on the vertical axis, and the *Vibrio* ORF 7 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity.

Figure 11 demonstrates sequence homology of ORF 8 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 8 is depicted on the vertical axis, and the *Vibrio* ORF 8 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity.

Figure 12 demonstrates sequence homology of ORF 9 of *Shewanella putrefaciens* and *Vibrio marinus*. The *Shewanella* ORF 9 is depicted on the vertical axis, and the *Vibrio* ORF 9 is depicted on the horizontal axis. Lines indicate regions of the proteins that have a 60% identity.

Figure 13 is a depiction of various complementation experiments, and resulting PUFA production. On the right, is shown the longest PUFA made in the *E. coli* strain containing the *Vibrio* and *Shewanella* genes depicted on the left. The hollow boxes indicate ORFs from *Shewanella*. The solid boxes indicate ORFs from *Vibrio*.

Figure 14 is a chromatogram showing fatty acid production from complementation of pEPAD8 from *Shewanella* (deletion ORF 8) with ORF 8 from *Shewanella*, in *E. coli* Fad E-. The chromatogram presents an EPA (20:5) peak.

Figure 15 is a chromatogram showing fatty acid production from complementation of pEPAD8 from *Shewanella* (deletion ORF 8) with ORF 8 from *Vibrio marinus*, in *E. coli* Fad E-. The chromatogram presents EPA (20:5) and DHA (22:6) peaks.

Figure 16 is a table of PUFA values from the ORF 8 complementation experiment, the chromatogram of which is shown in Figure 15.

Figure 17 is a plasmid map showing the elements of pCGN7770.

Figure 18 is a plasmid map showing the elements of pCGN8535.

Figure 19 is a plasmid map showing the elements of pCGN8537.

Figure 20 is a plasmid map showing the elements of pCGN8525.

Figure 21 is a comparison of the *Shewanella* ORFs as defined by Yazawa (1996) supra, and those disclosed in Figure 4. When a protein starting at the leucine (TTG) codon at nucleotides 9157-9155 and ending at the stop codon at nucleotides 8185-8183 is expressed under control of a heterologous promoter in an *E. coli* strain containing the entire PKS-like

cluster except ORF 3, the recombinant cells do produce EPA. Thus, the published protein sequence is likely to be wrong, and the coding sequence for the protein may start at the TTG codon at nucleotides 9157-9155 or the TTG codon at nucleotides 9172-9170. This information is critical to the expression of a functional PKS-like cluster heterologous system.

5 Figure 22 is a plasmid map showing the elements of pCGN8560.

Figure 23 is plasmid map showing the elements of pCGN8556.

Figure 24 shows the translated DNA sequence (SEQ ID NO:14) upstream of the published ORF 3 and the corresponding amino acids for which they code (SEQ ID NO:15). The ATG start codon at position 9016 is the start codon for the protein described by Yazawa *et al*  
10 (1996) *supra*. The other arrows depict TTG or ATT codons that can also serve as start codons in bacteria. When ORF 3 is started from the published ATG codon at 9016, the protein is not functional in making EPA. When ORF 3 is initiated at the TTG codon at position 9157, the protein is capable of facilitating EPA synthesis.

Figure 25 shows the PCR product (SEQ ID NO:16) for SS9 Photobacter using primers in  
15 Example 1.

Figure 26 shows probe sequences (SEQ ID NOS:17-31) resulting from PCR with primers presented in Example 1.

Figure 27 shows the nucleotide sequence of *Schizochytrium* EST clones A. LIB 3033-047-B5, LIB3033-046-E6 and a bridging PCR product have now been assembled into a partial  
20 cDNA sequence (ORF6 homolog), B. LIB3033-046-D2 (hglc/ORF7/ORF8/ORF9 homolog), C. LIB81-015-D5, LIB81-042-B9 and a bridging PCR product have now been assembled into a partial cDNA sequence (ORF8/ORF9 homolog).

Figure 28 shows a schematic of the similarities between *Shewanella* PKS sequences and *Schizochytrium* sequences.

25 Figure 29 shows the amino acid sequences inferred from *Schizochytrium* EST clones A. ORF6 homolog, B. hglc/ORF7/ORF8/ORF9 homolog, C. ORF8/ORF9 homolog.

### **DESCRIPTION OF THE PREFERRED EMBODIMENTS**

In accordance with the subject invention, novel DNA sequences, DNA constructs and  
30 methods are provided, which include some or all of the polyketide-like synthesis (PKS-like) pathway genes from *Shewanella*, *Vibrio*, *Schizochytrium* or other microorganisms, for modifying the poly-unsaturated long chain fatty acid content of host cells, particularly host plant cells. The present invention demonstrates that EPA synthesis genes in *Shewanella putrefaciens* constitute a polyketide-like synthesis pathway. Functions are ascribed to the *Shewanella*,  
35 *Schizochytrium* and *Vibrio* genes and methods are provided for the production of EPA and DHA in host cells. The method includes the step of transforming cells with an expression cassette comprising a DNA encoding a polypeptide capable of increasing the amount of one or more

PUFA in the host cell. Desirably, integration constructs are prepared which provide for integration of the expression cassette into the genome of a host cell. Host cells are manipulated to express a sense or antisense DNA encoding a polypeptide(s) that has PKS-like gene activity. By "PKS-like gene" is intended a polypeptide which is responsible for any one or more of the functions of a PKS-like activity of interest. By "polypeptide" is meant any chain of amino acids, regardless of length or post-translational modification, for example, glycosylation or phosphorylation. Depending upon the nature of the host cell, the substrate(s) for the expressed enzyme may be produced by the host cell or may be exogenously supplied. Of particular interest is the selective control of PUFA production in plant tissues and/or plant parts such as leaves, roots, fruits and seeds. The invention can be used to synthesize EPA, DHA, and other related PUFAs in host cells.

There are many advantages to transgenic production of PUFAs. As an example, in transgenic *E. coli* as in *Shewanella*, EPA accumulates in the phospholipid fraction, specifically in the *sn*-2 position. It may be possible to produce a structured lipid in a desired host cell which differs substantially from that produced in either *Shewanella* or *E. coli*. Additionally transgenic production of PUFAs in particular host cells offers several advantages over purification from natural sources such as fish or plants. In transgenic plants, by utilizing a PKS-like system, fatty acid synthesis of PUFAs is achieved in the cytoplasm by a system which produces the PUFAs through *de novo* production of the fatty acids utilizing malonyl Co-A and acetyl Co-A as substrates. In this fashion, potential problems, such as those associated with substrate competition and diversion of normal products of fatty acid synthesis in a host to PUFA production, are avoided.

Production of fatty acids from recombinant plants provides the ability to alter the naturally occurring plant fatty acid profile by providing new synthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs. Production of fatty acids in transgenic plants also offers the advantage that expression of PKS-like genes in particular tissues and/or plant parts means that greatly increased levels of desired PUFAs in those tissues and/or parts can be achieved, making recovery from those tissues more economical. Expression in a plant tissue and/or plant part presents certain efficiencies, particularly where the tissue or part is one which is easily harvested, such as seed, leaves, fruits, flowers, roots, etc. For example, the desired PUFAs can be expressed in seed; methods of isolating seed oils are well established. In addition to providing a source for purification of desired PUFAs, seed oil components can be manipulated through expression of PKS-like genes, either alone or in combination with other genes such as elongases, to provide seed oils having a particular PUFA profile in concentrated form. The concentrated seed oils then can be added to animal milks and/or synthetic or

semisynthetic milks to serve as infant formulas where human nursing is impossible or undesired, or in cases of malnourishment or disease in both adults and infants.

Transgenic microbial production of fatty acids offers the advantages that many microbes are known with greatly simplified oil compositions as compared with those of higher organisms, making purification of desired components easier. Microbial production is not subject to  
5 fluctuations caused by external variables such as weather and food supply. Microbially produced oil is substantially free of contamination by environmental pollutants. Additionally, microbes can provide PUFAs in particular forms which may have specific uses. For example, *Spirulina* can provide PUFAs predominantly at the first and third positions of triglycerides;  
10 digestion by pancreatic lipases preferentially releases fatty acids from these positions. Following human or animal ingestion of triglycerides derived from *Spirulina*, these PUFAs are released by pancreatic lipases as free fatty acids and thus are directly available, for example, for infant brain development. Additionally, microbial oil production can be manipulated by controlling culture conditions, notably by providing particular substrates for microbially  
15 expressed enzymes, or by addition of compounds which suppress undesired biochemical pathways. In addition to these advantages, production of fatty acids from recombinant microbes provides the ability to alter the naturally occurring microbial fatty acid profile by providing new synthetic pathways in the host or by suppressing undesired pathways, thereby increasing levels of desired PUFAs, or conjugated forms thereof, and decreasing levels of undesired PUFAs.

20 Production of fatty acids in animals also presents several advantages. Expression of desaturase genes in animals can produce greatly increased levels of desired PUFAs in animal tissues, making recovery from those tissues more economical. For example, where the desired PUFAs are expressed in the breast milk of animals, methods of isolating PUFAs from animal milk are well established. In addition to providing a source for purification of desired PUFAs,  
25 animal breast milk can be manipulated through expression of desaturase genes, either alone or in combination with other human genes, to provide animal milks with a PUFA composition substantially similar to human breast milk during the different stages of infant development. Humanized animal milks could serve as infant formulas where human nursing is impossible or undesired, or in the cases of malnourishment or disease.

30 DNAs encoding desired PKS-like genes can be identified in a variety of ways. In one method, a source of a desired PKS-like gene, for example genomic libraries from a *Shewanella*, *Schizochytrium* or *Vibrio* spp., is screened with detectable enzymatically- or chemically-synthesized probes. Sources of ORFs having PKS-like genes are those organisms which produce a desired PUFA, including DHA-producing or EPA-producing deep sea bacteria growing  
35 preferentially under high pressure or at relatively low temperature. Microorganisms such as *Shewanella* which produce EPA or DHA also can be used as a source of PKS-like genes. The probes can be made from DNA, RNA, or non-naturally occurring nucleotides, or mixtures



thereof. Probes can be enzymatically synthesized from DNAs of known PKS-like genes for normal or reduced-stringency hybridization methods. For discussions of nucleic acid probe design and annealing conditions, see, for example, Sambrook *et al*, *Molecular Cloning: A Laboratory Manual* (2<sup>nd</sup> ed.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) or *Current Protocols in Molecular Biology*, F. Ausubel *et al*, ed., Greene Publishing and Wiley-Interscience, New York (1987), each of which is incorporated herein by reference. Techniques for manipulation of nucleic acids encoding PUFA enzymes such as subcloning nucleic acid sequences encoding polypeptides into expression vectors, labelling probes, DNA hybridization, and the like are described generally in Sambrook, *supra*.

Oligonucleotide probes also can be used to screen sources and can be based on sequences of known PKS-like genes, including sequences conserved among known PKS-like genes, or on peptide sequences obtained from a desired purified protein. Oligonucleotide probes based on amino acid sequences can be degenerate to encompass the degeneracy of the genetic code, or can be biased in favor of the preferred codons of the source organism. Alternatively, a desired protein can be entirely sequenced and total synthesis of a DNA encoding that polypeptide performed.

Once the desired DNA has been isolated, it can be sequenced by known methods. It is recognized in the art that such methods are subject to errors, such that multiple sequencing of the same region is routine and is still expected to lead to measurable rates of mistakes in the resulting deduced sequence, particularly in regions having repeated domains, extensive secondary structure, or unusual base compositions, such as regions with high GC base content. When discrepancies arise, resequencing can be done and can employ special methods. Special methods can include altering sequencing conditions by using: different temperatures; different enzymes; proteins which alter the ability of oligonucleotides to form higher order structures; altered nucleotides such as ITP or methylated dGTP; different gel compositions, for example adding formamide; different primers or primers located at different distances from the problem region; or different templates such as single stranded DNAs. Sequencing of mRNA can also be employed.

For the most part, some or all of the coding sequences for the polypeptides having PKS-like gene activity are from a natural source. In some situations, however, it is desirable to modify all or a portion of the codons, for example, to enhance expression, by employing host preferred codons. Host preferred codons can be determined from the codons of highest frequency in the proteins expressed in the largest amount in a particular host species of interest. Thus, the coding sequence for a polypeptide having PKS-like gene activity can be synthesized in whole or in part. All or portions of the DNA also can be synthesized to remove any destabilizing sequences or regions of secondary structure which would be present in the transcribed mRNA. All or portions of the DNA also can be synthesized to alter the base

composition to one more preferable to the desired host cell. Methods for synthesizing sequences and bringing sequences together are well established in the literature. *In vitro* mutagenesis and selection, site-directed mutagenesis, or other means can be employed to obtain mutations of naturally occurring PKS-like genes to produce a polypeptide having PKS-like gene activity *in vivo* with more desirable physical and kinetic parameters for function in the host cell, such as a longer half-life or a higher rate of production of a desired polyunsaturated fatty acid.

Of particular interest are the *Shewanella putrefaciens* ORFs and the corresponding ORFs of *Vibrio marinus* and *Schizochytrium*. The *Shewanella putrefaciens* PKS-like genes can be expressed in transgenic plants to effect biosynthesis of EPA. Other DNAs which are substantially identical in sequence to the *Shewanella putrefaciens* PKS-like genes, or which encode polypeptides which are substantially similar to PKS-like genes of *Shewanella putrefaciens* can be used, such as those identified from *Vibrio marinus* or *Schizochytrium*. By substantially identical in sequence is intended an amino acid sequence or nucleic acid sequence exhibiting in order of increasing preference at least 60%, 80%, 90% or 95% homology to the DNA sequence of the *Shewanella putrefaciens* PKS-like genes or nucleic acid sequences encoding the amino acid sequences for such genes. For polypeptides, the length of comparison sequences generally is at least 16 amino acids, preferably at least 20 amino acids, and most preferably 35 amino acids. For nucleic acids, the length of comparison sequences generally is at least 50 nucleotides, preferably at least 60 nucleotides, and more preferably at least 75 nucleotides, and most preferably, 110 nucleotides.

Homology typically is measured using sequence analysis software, for example, the Sequence Analysis software package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wisconsin 53705, MEGAlign (DNASar, Inc., 1228 S. Park St., Madison, Wisconsin 53715), and MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, California 95008). BLAST (National Center for Biotechnology Information (WCBI) [www.ncbi.nlm.gov](http://www.ncbi.nlm.gov); FASTA (Pearson and Lipman, *Science* (1985) 227:1435-1446). Such software matches similar sequences by assigning degrees of homology to various substitutions, deletions, and other modifications. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine and leucine; aspartic acid, glutamic acid, asparagine, and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine. Substitutions may also be made on the basis of conserved hydrophobicity or hydrophilicity (Kyte and Doolittle, *J. Mol. Biol.* (1982) 157: 105-132), or on the basis of the ability to assume similar polypeptide secondary structure (Chou and Fasman, *Adv. Enzymol.* (1978) 47: 45-148, 1978). A related protein to the probing sequence is identified when  $p \geq 0.01$ , preferably  $p \geq 10^{-7}$  or  $10^{-8}$ .

Encompassed by the present invention are related PKS-like genes from the same or other organisms. Such related PKS-like genes include variants of the disclosed PKS-like ORFs that occur naturally within the same or different species of *Shewanella*, as well as homologues of the disclosed PKS-like genes from other species and evolutionarily related proteins having analogous function and activity. Also included are PKS-like genes which, although not substantially identical to the *Shewanella putrefaciens* PKS-like genes, operate in a similar fashion to produce PUFAs as part of a PKS-like system. Related PKS-like genes can be identified by their ability to function substantially the same as the disclosed PKS-like genes; that is, they can be substituted for corresponding ORFs of *Shewanella*, *Schizochytrium* or *Vibrio* and still effectively produce EPA or DHA. Related PKS-like genes also can be identified by screening sequence databases for sequences homologous to the disclosed PKS-like genes, by hybridization of a probe based on the disclosed PKS-like genes to a library constructed from the source organism, or by RT-PCR using mRNA from the source organism and primers based on the disclosed PKS-like gene. Thus, the phrase "PKS-like genes" refers not only to the nucleotide sequences disclosed herein, but also to other nucleic acids that are allelic or species variants of these nucleotide sequences. It is also understood that these terms include nonnatural mutations introduced by deliberate mutation using recombinant technology such as single site mutation or by excising short sections of DNA open reading frames coding for PUFA enzymes or by substituting new codons or adding new codons. Such minor alterations substantially maintain the immunoidentity of the original expression product and/or its biological activity. The biological properties of the altered PUFA enzymes can be determined by expressing the enzymes in an appropriate cell line and by determining the ability of the enzymes to synthesize PUFAs. Particular enzyme modifications considered minor would include substitution of amino acids of similar chemical properties, e.g., glutamic acid for aspartic acid or glutamine for asparagine.

When utilizing a PUFA PKS-like system from another organism, the regions of a PKS-like gene polypeptide important for PKS-like gene activity can be determined through routine mutagenesis, expression of the resulting mutant polypeptides and determination of their activities. The coding region for the mutants can include deletions, insertions and point mutations, or combinations thereof. A typical functional analysis begins with deletion mutagenesis to determine the N- and C-terminal limits of the protein necessary for function, and then internal deletions, insertions or point mutants are made in the open ready frame to further determine regions necessary for function. Other techniques such as cassette mutagenesis or total synthesis also can be used. Deletion mutagenesis is accomplished, for example, by using exonucleases to sequentially remove the 5' or 3' coding regions. Kits are available for such techniques. After deletion, the coding region is completed by ligating oligonucleotides containing start or stop codons to the deleted coding region after 5' or 3' deletion, respectively.

Alternatively, oligonucleotides encoding start or stop codons are inserted into the coding region by a variety of methods including site-directed mutagenesis, mutagenic PCR or by ligation onto DNA digested at existing restriction sites. Internal deletions can similarly be made through a variety of methods including the use of existing restriction sites in the DNA, by use of mutagenic primers via site directed mutagenesis or mutagenic PCR. Insertions are made through methods such as linker-scanning mutagenesis, site-directed mutagenesis or mutagenic PCR. Point mutations are made through techniques such as site-directed mutagenesis or mutagenic PCR.

Chemical mutagenesis also can be used for identifying regions of a PKS-like gene polypeptide important for activity. A mutated construct is expressed, and the ability of the resulting altered protein to function as a PKS-like gene is assayed. Such structure-function analysis can determine which regions may be deleted, which regions tolerate insertions, and which point mutations allow the mutant protein to function in substantially the same way as the native PKS-like gene. All such mutant proteins and nucleotide sequences encoding them are within the scope of the present invention. EPA is produced in *Shewanella* as the product of a PKS-like system, such that the EPA genes encode components of this system. In *Vibrio*, DHA is produced by a similar system. The enzymes which synthesize these fatty acids are encoded by a cluster of genes which are distinct from the fatty acid synthesis genes encoding the enzymes involved in synthesis of the C16 and C18 fatty acids typically found in bacteria and in plants. As the *Shewanella* EPA genes represent a PKS-like gene cluster, EPA production is, at least to some extent, independent of the typical bacterial type II FAS system. Thus, production of EPA in the cytoplasm of plant cells can be achieved by expression of the PKS-like pathway genes in plant cells under the control of appropriate plant regulatory signals.

EPA production in *E. coli* transformed with the *Shewanella* EPA genes proceeds during anaerobic growth, indicating that O<sub>2</sub>-dependent desaturase reactions are not involved. Analyses of the proteins encoded by the ORFs essential for EPA production reveals the presence of domain structures characteristic of PKS-like systems. Fig. 2A shows a summary of the domains, motifs, and also key homologies detected by "BLAST" data bank searches. Because EPA is different from many of the other substances produced by PKS-like pathways, i.e., it contains 5, *cis* double bonds, spaced at 3 carbon intervals along the molecule, a PKS-like system for synthesis of EPA is not expected.

Further, BLAST searches using the domains present in the *Shewanella* EPA ORFs reveal that several are related to proteins encoded by a PKS-like gene cluster found in Anabaena. The structure of that region of the Anabaena chromosome is shown in Fig. 2F. The Anabaena PKS-like genes have been linked to the synthesis of a long-chain (C26), hydroxy-fatty acid found in a glycolipid layer of heterocysts. The EPA protein domains with homology to the Anabaena proteins are indicated in Fig. 2F.

ORF 6 of *Shewanella* contains a KAS domain which includes an active site motif (DXAC\*), SEQ ID NO:32, as well as a "GFGG", SEQ ID NO:33, motif which is present at the end of many Type II KAS proteins (see Fig. 2A). Extended motifs are present but not shown here. Next is a malonyl-CoA:ACP acyl transferase (AT) domain. Sequences near the active site motif (GHS\*XG), SEQ ID NO:34, suggest it transfers malonate rather than methylmalonate, i.e., it resembles the acetate-like ATs. Following a linker region, there is a cluster of 6 repeating domains, each ~100 amino acids in length, which are homologous to PKS-like ACP sequences. Each contains a pantetheine binding site motif (LGXDS\*(L/I)), SEQ ID NOS:35 and 36. The presence of 6 such ACP domains has not been observed previously in fatty acid synthases (FAS) or PKS-like systems. Near the end of the protein is a region which shows homology to  $\beta$ -keto-ACP reductases (KR). It contains a pyridine nucleotide binding site motif "GXGXX(G/A/P)", SEQ ID NOS:37, 38 and 39.

The *Shewanella* ORF 8 begins with a KAS domain, including active site and ending motifs (Fig. 2C). The best match in the data banks is with the Anabeana HglD. There is also a domain which has sequence homology to the N-terminal one half of the Anabeana HglC. This region also shows weak homology to KAS proteins although it lacks the active site and ending motifs. It has the characteristics of the so-called chain length factors (CLF) of Type II PKS-like systems. ORF 8 appears to direct the production of EPA versus DHA by the PKS-like system. ORF 8 also has two domains with homology to  $\beta$ -hydroxyacyl-ACP dehydrases (DH). The best match for both domains is with *E. coli* FabA, a bi-functional enzyme which carries out both the dehydrase reaction and an isomerization (*trans* to *cis*) of the resulting double bond. The first DH domain contains both the active site histidine (H) and an adjacent cysteine (C) implicated in FabA catalysis. The second DH domain has the active site H but lacks the adjacent C (Fig. 2C). Blast searches with the second DH domain also show matches to FabZ, a second *E. coli* DH, which does not possess isomerase activity.

The N-terminal half of ORF 7 (Fig. 2B) has no significant matches in the data banks. The best match of the C-terminal half is with a C-terminal portion of the Anabeana HglC. This domain contains an acyl-transferase (AT) motif (GX SXG), SEQ ID NO:40. Comparison of the extended active site sequences, based on the crystal structure of the *E. coli* malonyl-CoA:ACP AT, reveals that ORF 7 lacks two residues essential for exclusion of water from the active site (*E. coli* nomenclature; Q11 and R117). These data suggest that ORF 7 may function as a thioesterase.

ORF 9 (Fig. 2D) is homologous to an ORF of unknown function in the Anabeana Hgl cluster. It also exhibits a very weak homology to NIFA, a regulatory protein in nitrogen fixing bacteria. A regulatory role for the ORF 9 protein has not been excluded. ORF 3 (Fig. 2E) is homologous to the Anabeana HetI as well as EntD from *E. coli* and Sfp of *Bacillus*. Recently, a new enzyme family of phosphopantetheinyl transferases has been identified that includes HetI,

EntD and Sfp (Lamblot RH, *et al.* (1996) A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & Biology*, Vol 3, #11, 923-936 ). The data of Fig. 3 demonstrates that the presence of ORF 3 is required for addition of  $\beta$ -alanine (i.e. pantetheine) to the ORF 6 protein. Thus, ORF 3 encodes the phosphopantetheinyl transferase specific for the ORF 6 ACP domains. (See, Haydock SF *et al.* (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases, *FEBS Lett.*, 374, 246-248). Malonate is the source of the carbons utilized in the extension reactions of EPA synthesis. Additionally, malonyl-CoA rather than malonyl-ACP is the AT substrate, i.e., the AT region of ORF 6 uses malonyl Co-A.

Once the DNA sequences encoding the PKS-like genes of an organism responsible for PUFA production have been obtained, they are placed in a vector capable of replication in a host cell, or propagated *in vitro* by means of techniques such as PCR or long PCR. Replicating vectors can include plasmids, phage, viruses, cosmids and the like. Desirable vectors include those useful for mutagenesis of the gene of interest or for expression of the gene of interest in host cells. A PUFA synthesis enzyme or a homologous protein can be expressed in a variety of recombinantly engineered cells. Numerous expression systems are available for expression of DNA encoding a PUFA enzyme. The expression of natural or synthetic nucleic acids encoding PUFA enzyme is typically achieved by operably linking the DNA to a promoter (which is either constitutive or inducible) within an expression vector. By expression vector is meant a DNA molecule, linear or circular, that comprises a segment encoding a PUFA enzyme, operably linked to additional segments that provide for its transcription. Such additional segments include promoter and terminator sequences. An expression vector also may include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, etc. Expression vectors generally are derived from plasmid or viral DNA, and can contain elements of both. The term "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, for example, transcription initiates in the promoter and proceeds through the coding segment to the terminator. See Sambrook *et al*, *supra*.

The technique of long PCR has made *in vitro* propagation of large constructs possible, so that modifications to the gene of interest, such as mutagenesis or addition of expression signals, and propagation of the resulting constructs can occur entirely *in vitro* without the use of a replicating vector or a host cell. *In vitro* expression can be accomplished, for example, by placing the coding region for the desaturase polypeptide in an expression vector designed for *in vitro* use and adding rabbit reticulocyte lysate and cofactors; labeled amino acids can be incorporated if desired. Such *in vitro* expression vectors may provide some or all of the expression signals necessary in the system used. These methods are well known in the art and the components of the system are commercially available. The reaction mixture can then be

assayed directly for PKS-like enzymes for example by determining their activity, or the synthesized enzyme can be purified and then assayed.

Expression in a host cell can be accomplished in a transient or stable fashion. Transient expression can occur from introduced constructs which contain expression signals functional in the host cell, but which constructs do not replicate and rarely integrate in the host cell, or where the host cell is not proliferating. Transient expression also can be accomplished by inducing the activity of a regulatable promoter operably linked to the gene of interest, although such inducible systems frequently exhibit a low basal level of expression. Stable expression can be achieved by introduction of a nucleic acid construct that can integrate into the host genome or that autonomously replicates in the host cell. Stable expression of the gene of interest can be selected for through the use of a selectable marker located on or transfected with the expression construct, followed by selection for cells expressing the marker. When stable expression results from integration, integration of constructs can occur randomly within the host genome or can be targeted through the use of constructs containing regions of homology with the host genome sufficient to target recombination with the host locus. Where constructs are targeted to an endogenous locus, all or some of the transcriptional and translational regulatory regions can be provided by the endogenous locus. To achieve expression in a host cell, the transformed DNA is operably associated with transcriptional and translational initiation and termination regulatory regions that are functional in the host cell.

Transcriptional and translational initiation and termination regions are derived from a variety of nonexclusive sources, including the DNA to be expressed, genes known or suspected to be capable of expression in the desired system, expression vectors, chemical synthesis. The termination region can be derived from the 3' region of the gene from which the initiation region was obtained or from a different gene. A large number of termination regions are known to and have been found to be satisfactory in a variety of hosts from the same and different genera and species. The termination region usually is selected more as a matter of convenience rather than because of any particular property. When expressing more than one PKS-like ORF in the same cell, appropriate regulatory regions and expression methods should be used. Introduced genes can be propagated in the host cell through use of replicating vectors or by integration into the host genome. Where two or more genes are expressed from separate replicating vectors, it is desirable that each vector has a different means of replication. Each introduced construct, whether integrated or not, should have a different means of selection and should lack homology to the other constructs to maintain stable expression and prevent reassortment of elements among constructs. Judicious choices of regulatory regions, selection means and method of propagation of the introduced construct can be experimentally determined so that all introduced genes are expressed at the necessary levels to provide for synthesis of the desired products.

A variety of procaryotic expression systems can be used to express PUFA enzyme. Expression vectors can be constructed which contain a promoter to direct transcription, a ribosome binding site, and a transcriptional terminator. Examples of regulatory regions suitable for this purpose in *E. coli* are the promoter and operator region of the *E. coli* tryptophan biosynthetic pathway as described by Yanofsky (1984) *J. Bacteriol.*, 158:1018-1024 and the leftward promoter of phage lambda ( $P_{\lambda}$ ) as described by Herskowitz and Hagen, (1980) *Ann. Rev. Genet.*, 14:399-445. The inclusion of selection markers in DNA vectors transformed in *E. coli* is also useful. Examples of such markers include genes specifying resistance to ampicillin, tetracycline, or chloramphenicol. Vectors used for expressing foreign genes in bacterial hosts generally will contain a selectable marker, such as a gene for antibiotic resistance, and a promoter which functions in the host cell. Plasmids useful for transforming bacteria include pBR322 (Bolivar, *et al.*, (1977) *Gene* 2:95-113), the pUC plasmids (Messing, (1983) *Meth. Enzymol.* 101:20-77, Vieira and Messing, (1982) *Gene* 19:259-268), pCQV2 (Queen, *ibid.*), and derivatives thereof. Plasmids may contain both viral and bacterial elements. Methods for the recovery of the proteins in biologically active form are discussed in U.S. Patent Nos. 4,966,963 and 4,999,422, which are incorporated herein by reference. See Sambrook, *et al* for a description of other prokaryotic expression systems.

For expression in eukaryotes, host cells for use in practicing the present invention include mammalian, avian, plant, insect, and fungal cells. As an example, for plants, the choice of a promoter will depend in part upon whether constitutive or inducible expression is desired and whether it is desirable to produce the PUFAs at a particular stage of plant development and/or in a particular tissue. Considerations for choosing a specific tissue and/or developmental stage for expression of the ORFs may depend on competing substrates or the ability of the host cell to tolerate expression of a particular PUFA. Expression can be targeted to a particular location within a host plant such as seed, leaves, fruits, flowers, and roots, by using specific regulatory sequences, such as those described in USPN 5,463,174, USPN 4,943,674, USPN 5,106,739, USPN 5,175,095, USPN 5,420,034, USPN 5,188,958, and USPN 5,589,379. Where the host cell is a yeast, transcription and translational regions functional in yeast cells are provided, particularly from the host species. The transcriptional initiation regulatory regions can be obtained, for example from genes in the glycolytic pathway, such as alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase (GPD), phosphoglucoisomerase, phosphoglycerate kinase, etc. or regulatable genes such as acid phosphatase, lactase, metallothionein, glucoamylase, etc. Any one of a number of regulatory sequences can be used in a particular situation, depending upon whether constitutive or induced transcription is desired, the particular efficiency of the promoter in conjunction with the open-reading frame of interest, the ability to join a strong promoter with a control region from a different promoter which allows for inducible transcription, ease of construction, and the like. Of particular interest are promoters



which are activated in the presence of galactose. Galactose-inducible promoters (GAL1, GAL7, and GAL10) have been extensively utilized for high level and regulated expression of protein in yeast (Lue *et al*, (1987) *Mol. Cell. Biol.* 7:3446; Johnston, (1987) *Microbiol. Rev.* 51:458).

Transcription from the GAL promoters is activated by the GAL4 protein, which binds to the promoter region and activates transcription when galactose is present. In the absence of galactose, the antagonist GAL80 binds to GAL4 and prevents GAL4 from activating transcription. Addition of galactose prevents GAL80 from inhibiting activation by GAL4.

Preferably, the termination region is derived from a yeast gene, particularly *Saccharomyces*, *Schizosaccharomyces*, *Candida* or *Kluyveromyces*. The 3' regions of two mammalian genes,  $\gamma$  interferon and  $\alpha 2$  interferon, are also known to function in yeast.

Nucleotide sequences surrounding the translational initiation codon ATG have been found to affect expression in yeast cells. If the desired polypeptide is poorly expressed in yeast, the nucleotide sequences of exogenous genes can be modified to include an efficient yeast translation initiation sequence to obtain optimal gene expression. For expression in *Saccharomyces*, this can be done by site-directed mutagenesis of an inefficiently expressed gene by fusing it in-frame to an endogenous *Saccharomyces* gene, preferably a highly expressed gene, such as the lactase gene.

As an alternative to expressing the PKS-like genes in the plant cell cytoplasm, is to target the enzymes to the chloroplast. One method to target proteins to the chloroplast entails use of leader peptides attached to the N-termini of the proteins. Commonly used leader peptides are derived from the small subunit of plant ribulose bis phosphate carboxylase. Leader sequences from other chloroplast proteins may also be used. Another method for targeting proteins to the chloroplast is to transform the chloroplast genome (Stable transformation of chloroplasts of *Chlamydomonas reinhardtii* (1 green alga) using bombardment of recipient cells with high-velocity tungsten microprojectiles coated with foreign DNA has been described. See, for example, Blowers *et al Plant Cell* (1989) 1:123-132 and Debuchy *et al EMBO J* (1989) 8:2803-2809. The transformation technique, using tungsten microprojectiles, is described by Kline *et al, Nature* (London) (1987) 327:70-73). The most common method of transforming chloroplasts involves using biolistic techniques, but other techniques developed for the purpose may also be used. (Methods for targeting foreign gene products into chloroplasts (Shrier *et al EMBO J.* (1985) 4:25-32) or mitochondria (Boutry *et al, supra*) have been described. See also Tomai *et al Gen. Biol. Chem.* (1988) 263:15104-15109 and US Patent No. 4,940,835 for the use of transit peptides for translocating nuclear gene products into the chloroplast. Methods for directing the transport of proteins to the chloroplast are reviewed in Kenauf *TIBTECH* (1987) 5:40-47.

For producing PUFAs in avian species and cells, gene transfer can be performed by introducing a nucleic acid sequence encoding a PUFA enzyme into the cells following procedures known in the art. If a transgenic animal is desired, pluripotent stem cells of embryos

can be provided with a vector carrying a PUFA enzyme encoding transgene and developed into adult animal (USPN 5,162,215; Ono *et al.* (1996) *Comparative Biochemistry and Physiology A* 113(3):287-292; WO 9612793; WO 9606160). In most cases, the transgene is modified to express high levels of the PKS-like enzymes in order to increase production of PUFAs. The transgenes can be modified, for example, by providing transcriptional and/or translational regulatory regions that function in avian cells, such as promoters which direct expression in particular tissues and egg parts such as yolk. The gene regulatory regions can be obtained from a variety of sources, including chicken anemia or avian leukosis viruses or avian genes such as a chicken ovalbumin gene.

Production of PUFAs in insect cells can be conducted using baculovirus expression vectors harboring PKS-like transgenes. Baculovirus expression vectors are available from several commercial sources such as Clontech. Methods for producing hybrid and transgenic strains of algae, such as marine algae, which contain and express a desaturase transgene also are provided. For example, transgenic marine algae can be prepared as described in USPN 5,426,040. As with the other expression systems described above, the timing, extent of expression and activity of the desaturase transgene can be regulated by fitting the polypeptide coding sequence with the appropriate transcriptional and translational regulatory regions selected for a particular use. Of particular interest are promoter regions which can be induced under preselected growth conditions. For example, introduction of temperature sensitive and/or metabolite responsive mutations into the desaturase transgene coding sequences, its regulatory regions, and/or the genome of cells into which the transgene is introduced can be used for this purpose.

The transformed host cell is grown under appropriate conditions adapted for a desired end result. For host cells grown in culture, the conditions are typically optimized to produce the greatest or most economical yield of PUFAs, which relates to the selected desaturase activity. Media conditions which may be optimized include: carbon source, nitrogen source, addition of substrate, final concentration of added substrate, form of substrate added, aerobic or anaerobic growth, growth temperature, inducing agent, induction temperature, growth phase at induction, growth phase at harvest, pH, density, and maintenance of selection. Microorganisms such as yeast, for example, are preferably grown using selected media of interest, which include yeast peptone broth (YPD) and minimal media (contains amino acids, yeast nitrogen base, and ammonium sulfate, and lacks a component for selection, for example uracil). Desirably, substrates to be added are first dissolved in ethanol. Where necessary, expression of the polypeptide of interest may be induced, for example by including or adding galactose to induce expression from a GAL promoter.

When increased expression of the PKS-like gene polypeptide in a host cell which expresses PUFA from a PKS-like system is desired, several methods can be employed.

Additional genes encoding the PKS-like gene polypeptide can be introduced into the host organism. Expression from the native PKS-like gene locus also can be increased through homologous recombination, for example by inserting a stronger promoter into the host genome to cause increased expression, by removing destabilizing sequences from either the mRNA or the encoded protein by deleting that information from the host genome, or by adding stabilizing sequences to the mRNA (*see* USPN 4,910,141 and USPN 5,500,365). Thus, the subject host will have at least have one copy of the expression construct and may have two or more, depending upon whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple copy numbers. Where the subject host is a yeast, four principal types of yeast plasmid vectors can be used: Yeast Integrating plasmids (YIps), Yeast Replicating plasmids (YRps), Yeast Centromere plasmids (YCps), and Yeast Episomal plasmids (YEps). YIps lack a yeast replication origin and must be propagated as integrated elements in the yeast genome. YRps have a chromosomally derived autonomously replicating sequence and are propagated as medium copy number (20 to 40), autonomously replicating, unstably segregating plasmids. YCps have both a replication origin and a centromere sequence and propagate as low copy number (10-20), autonomously replicating, stably segregating plasmids. YEps have an origin of replication from the yeast 2 $\mu$ m plasmid and are propagated as high copy number, autonomously replicating, irregularly segregating plasmids. The presence of the plasmids in yeast can be ensured by maintaining selection for a marker on the plasmid. Of particular interest are the yeast vectors pYES2 (a YEps plasmid available from Invitrogen, confers uracil prototrophy and a GAL1 galactose-inducible promoter for expression), and pYX424 (a YEps plasmid having a constitutive TP1 promoter and conferring leucine prototrophy; (Alber and Kawasaki (1982). *J. Mol. & Appl. Genetics* 1: 419).

The choice of a host cell is influenced in part by the desired PUFA profile of the transgenic cell, and the native profile of the host cell. Even where the host cell expresses PKS-like gene activity for one PUFA, expression of PKS-like genes of another PKS-like system can provide for production of a novel PUFA not produced by the host cell. In particular instances where expression of PKS-like gene activity is coupled with expression of an ORF 8 PKS-like gene of an organism which produces a different PUFA, it can be desirable that the host cell naturally have, or be mutated to have, low PKS-like gene activity for ORF 8. As an example, for production of EPA, the DNA sequence used encodes the polypeptide having PKS-like gene activity of an organism which produces EPA, while for production of DHA, the DNA sequences used are those from an organism which produces DHA. For use in a host cell which already expresses PKS-like gene activity it can be necessary to utilize an expression cassette which provides for overexpression of the desired PKS-like genes alone or with a construct to downregulate the activity of an existing ORF of the existing PKS-like system, such as by antisense or co-suppression. Similarly, a combination of ORFs derived from separate organisms

which produce the same or different PUFAs using PKS-like systems may be used. For instance, the ORF 8 of *Vibrio* directs the expression of DHA in a host cell, even when ORFs 3, 6, 7 and 9 are from *Shewanella*, which produce EPA when coupled to ORF 8 of *Shewanella*. Therefore, for production of eicosapentanoic acid (EPA), the expression cassettes used generally include one or more cassettes which include ORFs 3, 6, 7, 8 and 9 from a PUFA-producing organism such as the marine bacterium *Shewanella putrefaciens* (for EPA production) or *Vibrio marinus* (for DHA production). ORF 8 can be used for induction of DHA production, and ORF 8 of *Vibrio* can be used in conjunction with ORFs 3, 6, 7 and 9 of *Shewanella* to produce DHA. The organization and numbering scheme of the ORFs identified in the *Shewanella* gene cluster are shown in Fig 1A. Maps of several subclones referred to in this study are shown in Fig 1B. For expression of a PKS-like gene polypeptide, transcriptional and translational initiation and termination regions functional in the host cell are operably linked to the DNA encoding the PKS-like gene polypeptide.

Constructs comprising the PKS-like ORFs of interest can be introduced into a host cell by any of a variety of standard techniques, depending in part upon the type of host cell. These techniques include transfection, infection, bolistic impact, electroporation, microinjection, scraping, or any other method which introduces the gene of interest into the host cell (*see* USPN 4,743,548, USPN 4,795,855, USPN 5,068,193, USPN 5,188,958, USPN 5,463,174, USPN 5,565,346 and USPN 5,565,347). Methods of transformation which are used include lithium acetate transformation (*Methods in Enzymology*, (1991) 194:186-187). For convenience, a host cell which has been manipulated by any method to take up a DNA sequence or construct will be referred to as "transformed" or "recombinant" herein. The subject host will have at least one copy of the expression construct and may have two or more, depending upon whether the gene is integrated into the genome, amplified, or is present on an extrachromosomal element having multiple copy numbers.

For production of PUFAs, depending upon the host cell, the several polypeptides produced by pEPA, ORFs 3, 6, 7, 8 and 9, are introduced as individual expression constructs or can be combined into two or more cassettes which are introduced individually or co-transformed into a host cell. A standard transformation protocol is used. For plants, where less than all PKS-like genes required for PUFA synthesis have been inserted into a single plant, plants containing a complementing gene or genes can be crossed to obtain plants containing a full complement of PKS-like genes to synthesize a desired PUFA.

The PKS-like-mediated production of PUFAs can be performed in either prokaryotic or eukaryotic host cells. The cells can be cultured or formed as part or all of a host organism including an animal. Viruses and bacteriophage also can be used with appropriate cells in the production of PUFAs, particularly for gene transfer, cellular targeting and selection. Any type of plant cell can be used for host cells, including dicotyledonous plants, monocotyledonous plants,

and cereals. Of particular interest are crop plants such as *Brassica*, *Arabidopsis*, soybean, corn, and the like. Prokaryotic cells of interest include *Eschericia*, *Baccillus*, *Lactobaccillus*, *cyanobacteria* and the like. Eukaryotic cells include plant cells, mammalian cells such as those of lactating animals, avian cells such as of chickens, and other cells amenable to genetic manipulation including insect, fungal, and algae cells. Examples of host animals include mice, rats, rabbits, chickens, quail, turkeys, cattle, sheep, pigs, goats, yaks, etc., which are amenable to genetic manipulation and cloning for rapid expansion of a transgene expressing population. For animals, PKS-like transgenes can be adapted for expression in target organelles, tissues and body fluids through modification of the gene regulatory regions. Of particular interest is the production of PUFAs in the breast milk of the host animal.

Examples of host microorganisms include *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis*, or other yeast such as *Candida*, *Kluyveromyces* or other fungi, for example, filamentous fungi such as *Aspergillus*, *Neurospora*, *Penicillium*, etc. Desirable characteristics of a host microorganism are, for example, that it is genetically well characterized, can be used for high level expression of the product using ultra-high density fermentation, and is on the GRAS (generally recognized as safe) list since the proposed end product is intended for ingestion by humans. Of particular interest is use of a yeast, more particularly baker's yeast (*S. cerevisiae*), as a cell host in the subject invention. Strains of particular interest are SC334 (Mat  $\alpha$  pep4-3 prbl-1122 ura3-52 leu2-3, 112 reg1-501 gal1; (Hovland *et al* (1989) Gene 83:57-64); BJ1995 (Yeast Genetic Stock Centre, 1021 Donner Laboratory, Berkeley, CA 94720), INVSC1 (Mat  $\alpha$  hiw3 $\Delta$ 1 leu2 trp1-289 ura3-52 (Invitrogen, 1600 Faraday Ave., Carlsbad, CA 92008) and INVSC2 (Mat  $\alpha$  his3 $\Delta$ 200 ura3-167; (Invitrogen). Bacterial cells also may be used as hosts. This includes *E. coli*, which can be useful in fermentation processes. Alternatively, a host such as a *Lactobacillus* species can be used as a host for introducing the products of the PKS-like pathway into a product such as yogurt.

The transformed host cell can be identified by selection for a marker contained on the introduced construct. Alternatively, a separate marker construct can be introduced with the desired construct, as many transformation techniques introduce multiple DNA molecules into host cells. Typically, transformed hosts are selected for their ability to grow on selective media. Selective media can incorporate an antibiotic or lack a factor necessary for growth of the untransformed host, such as a nutrient or growth factor. An introduced marker gene therefor may confer antibiotic resistance, or encode an essential growth factor or enzyme, and permit growth on selective media when expressed in the transformed host cell. Desirably, resistance to kanamycin and the amino glycoside G418 are of particular interest (*see* USPN 5,034,322). For yeast transformants, any marker that functions in yeast can be used, such as the ability to grow on media lacking uracil, lencine, lysine or tryptophan.

Selection of a transformed host also can occur when the expressed marker protein can be detected, either directly or indirectly. The marker protein can be expressed alone or as a fusion to another protein. The marker protein can be one which is detected by its enzymatic activity; for example  $\beta$ -galactosidase can convert the substrate X-gal to a colored product, and luciferase  
5 can convert luciferin to a light-emitting product. The marker protein can be one which is detected by its light-producing or modifying characteristics; for example, the green fluorescent protein of *Aequorea victoria* fluoresces when illuminated with blue light. Antibodies can be used to detect the marker protein or a molecular tag on, for example, a protein of interest. Cells expressing the marker protein or tag can be selected, for example, visually, or by techniques  
10 such as FACS or panning using antibodies.

The PUFAs produced using the subject methods and compositions are found in the host plant tissue and/or plant part as free fatty acids and/or in conjugated forms such as acylglycerols, phospholipids, sulfolipids or glycolipids, and can be extracted from the host cell through a variety of means well-known in the art. Such means include extraction with organic solvents,  
15 sonication, supercritical fluid extraction using for example carbon dioxide, and physical means such as presses, or combinations thereof. Of particular interest is extraction with methanol and chloroform. Where appropriate, the aqueous layer can be acidified to protonate negatively charged moieties and thereby increase partitioning of desired products into the organic layer. After extraction, the organic solvents can be removed by evaporation under a stream of nitrogen.  
20 When isolated in conjugated forms, the products are enzymatically or chemically cleaved to release the free fatty acid or a less complex conjugate of interest, and are then subjected to further manipulations to produce a desired end product. Desirably, conjugated forms of fatty acids are cleaved with potassium hydroxide.

If further purification is necessary, standard methods can be employed. Such methods  
25 include extraction, treatment with urea, fractional crystallization, HPLC, fractional distillation, silica gel chromatography, high speed centrifugation or distillation, or combinations of these techniques. Protection of reactive groups, such as the acid or alkenyl groups, can be done at any step through known techniques, for example alkylation or iodination. Methods used include methylation of the fatty acids to produce methyl esters. Similarly, protecting groups can be  
30 removed at any step. Desirably, purification of fractions containing DHA and EPA is accomplished by treatment with urea and/or fractional distillation.

The uses of the subject invention are several. Probes based on the DNAs of the present invention find use in methods for isolating related molecules or in methods to detect organisms expressing PKS-like genes. When used as probes, the DNAs or oligonucleotides need to be  
35 detectable. This is usually accomplished by attaching a label either at an internal site, for example via incorporation of a modified residue, or at the 5' or 3' terminus. Such labels can be directly detectable, can bind to a secondary molecule that is detectably labeled, or can bind to an

unlabelled secondary molecule and a detectably labeled tertiary molecule; this process can be extended as long as is practicable to achieve a satisfactorily detectable signal without unacceptable levels of background signal. Secondary, tertiary, or bridging systems can include use of antibodies directed against any other molecule, including labels or other antibodies, or can involve any molecules which bind to each other, for example a biotin-streptavidin/avidin system. Detectable labels typically include radioactive isotopes, molecules which chemically or enzymatically produce or alter light, enzymes which produce detectable reaction products, magnetic molecules, fluorescent molecules or molecules whose fluorescence or light-emitting characteristics change upon binding. Examples of labelling methods can be found in USPN 5,011,770. Alternatively, the binding of target molecules can be directly detected by measuring the change in heat of solution on binding of a probe to a target via isothermal titration calorimetry, or by coating the probe or target on a surface and detecting the change in scattering of light from the surface produced by binding of a target or a probe, respectively, is done with the BIAcore system.

PUFAs produced by recombinant means find applications in a wide variety of areas. Supplementation of humans or animals with PUFAs in various forms can result in increased levels not only of the added PUFAs, but of their metabolic progeny as well. Complex regulatory mechanisms can make it desirable to combine various PUFAs, or to add different conjugates of PUFAs, in order to prevent, control or overcome such mechanisms to achieve the desired levels of specific PUFAs in an individual. In the present case, expression of PKS-like gene genes, or antisense PKS-like gene transcripts, can alter the levels of specific PUFAs, or derivatives thereof, found in plant parts and/or plant tissues. The PKS-like gene polypeptide coding region is expressed either by itself or with other genes, in order to produce tissues and/or plant parts containing higher proportions of desired PUFAs or containing a PUFA composition which more closely resembles that of human breast milk (Prieto *et al.*, PCT publication WO 95/24494) than does the unmodified tissues and/or plant parts.

PUFAs, or derivatives thereof, made by the disclosed method can be used as dietary supplements for patients undergoing intravenous feeding or for preventing or treating malnutrition. For dietary supplementation, the purified PUFAs, or derivatives thereof, can be incorporated into cooking oils, fats or margarines formulated so that in normal use the recipient receives a desired amount of PUFA. The PUFAs also can be incorporated into infant formulas, nutritional supplements or other food products, and find use as anti-inflammatory or cholesterol lowering agents.

Particular fatty acids such as EPA can be used to alter the composition of infant formulas to better replicate the PUFA composition of human breast milk. The predominant triglyceride in human milk is reported to be 1,3-di-oleoyl-2-palmitoyl, with 2-palmitoyl glycerides reported as better absorbed than 2-oleoyl or 2-lineoyl glycerides (*see* USPN 4,876,107). Typically, human

breast milk has a fatty acid profile comprising from about 0.15 % to about 0.36 % as DHA, from about 0.03 % to about 0.13 % as EPA, from about 0.30 % to about 0.88 % as ARA, from about 0.22 % to about 0.67 % as DGLA, and from about 0.27 % to about 1.04 % as GLA. A preferred ratio of GLA:DGLA:ARA in infant formulas is from about 1:1:4 to about 1:1:1, respectively.

- 5 Amounts of oils providing these ratios of PUFA can be determined without undue experimentation by one of skill in the art. PUFAs, or host cells containing them, also can be used as animal food supplements to alter an animal's tissue or milk fatty acid composition to one more desirable for human or animal consumption.

- 10 For pharmaceutical use (human or veterinary), the compositions generally are administered orally but can be administered by any route by which they may be successfully absorbed, e.g., parenterally (i.e. subcutaneously, intramuscularly or intravenously), rectally or vaginally or topically, for example, as a skin ointment or lotion. Where available, gelatin capsules are the preferred form of oral administration. Dietary supplementation as set forth above also can provide an oral route of administration. The unsaturated acids of the present invention can be administered in conjugated forms, or as salts, esters, amides or prodrugs of the fatty acids. Any pharmaceutically acceptable salt is encompassed by the present invention; especially preferred are the sodium, potassium or lithium salts. Also encompassed are the N-alkylpolyhydroxamine salts, such as N-methyl glucamine, described in PCT publication WO 96/33155. Preferred esters are the ethyl esters.

- 20 The PUFAs of the present invention can be administered alone or in combination with a pharmaceutically acceptable carrier or excipient. As solid salts, the PUFAs can also be administered in tablet form. For intravenous administration, the PUFAs or derivatives thereof can be incorporated into commercial formulations such as Intralipids. Where desired, the individual components of formulations can be individually provided in kit form, for single or multiple use. A typical dosage of a particular fatty acid is from 0.1 mg to 20 g, or even 100 g daily, and is preferably from 10 mg to 1, 2, 5 or 10 g daily as required, or molar equivalent amounts of derivative forms thereof. Parenteral nutrition compositions comprising from about 2 to about 30 weight percent fatty acids calculated as triglycerides are encompassed by the present invention. Other vitamins, and particularly fat-soluble vitamins such as vitamin A, D, E and L-carnitine optionally can be included. Where desired, a preservative such as a tocopherol can be added, typically at about 0.1% by weight.

The following examples are presented by way of illustration, not of limitation.



EXAMPLESExample 1The Identity of ORFs Derived from *Vibrio marinus*

5 Using polymerase chain reaction (PCR) with primers based on ORF 6 of *Shewanella* (Sp ORF 6) sequences (FW 5' primers CUACUACUACUACCAAGCT  
AAAGCACTTAACCGTG, SEQ ID NO:41, and CUACUACUACUACAGCGAAATG  
CTTATCAAG, SEQ ID NO:42, for *Vibrio* and SS9 respectively and 3' BW primers:  
CAUCAUCAUGCGACCAAAACCAAATGAGCTAATAC, SEQ ID NO:43, for both  
10 *Vibrio* and SS9) and genomic DNAs templates from *Vibrio* and a borophyllic *photobacter*  
producing EPA (provided by Dr. Bartlett, UC San Diego), resulted in PCR products of *ca.*400  
bases for *Vibrio marinus* (*Vibrio*) and *ca.*900 bases for SS9 presenting more than 75% homology  
with corresponding fragments of Sp ORF 6 (*see* Figure 25) as determined by direct counting of  
homologous amino acids.

15 A *Vibrio* cosmid library was then prepared and using the *Vibrio* ORF 6 PCR product as a  
probe (*see* Figure 26); clones containing at least ORF 6 were selected by colony hybridization.

Through additional sequences of the selected cosmids such as cosmid #9 and cosmid  
#21, a *Vibrio* cluster (Figure 5) with ORFs homologous to, and organized in the same sequential  
order (ORFs 6-9) as ORFs 6-9 of *Shewanella*, was obtained (Figure 7). The *Vibrio* ORFs from  
20 this sequence are found at 17394 to 36115 and comprehend ORFs 6-9.

Table*Vibrio* operon figures

	17394 to 25349	length = 7956 nt
25	25509 to 28157	length = 2649 nt
	28209 to 34262	length = 6054 nt
	34454 to 36115	length = 1662 nt

The ORF designations for the *Shewanella* genes are based on those disclosed in Figure 4, and  
30 differ from those published for the *Shewanella* cluster (Yazawa *et al*, USPN 5,683,898). For  
instance, ORF 3 of Figure 4 is read in the opposite direction from the other ORFs and is not  
disclosed in Yazawa *et al* USPN 5,683,898 (See Fig. 24) for comparison with Yazawa *et al*  
USPN 5,683,898.

Sequences homologous to ORF 3, were not found in the proximity of ORF 6 (17000  
35 bases upstream of ORF 6) or of ORF 9 (*ca.*4000 bases downstream of ORF 9). Motifs  
characteristic of phosphopantethenyl transferases (Lambalot *et al* (1996) *Current Biology* 3:923-

936) were absent from the *Vibrio* sequences screened for these motifs. In addition, there was no match to *Sp* ORF 3 derived probes in genomic digests of *Vibrio* and of SC2A *Shewanella* (another bacterium provided by the University of San Diego and also capable of producing EPA). Although ORF 3 may exist in *Vibrio*, its DNA may not be homologous to that of *Sp* ORF 3 and/or could be located in portions of the genome that were not sequenced.

Figure 6 provides the sequence of an approximately 19 kb *Vibrio* clone comprising ORFs 6-9. Figures 7 and 8 compare the gene cluster organizations of the PKS-like systems of *Vibrio marinus* and *Shewanella putrefaciens*. Figures 9 through 12 show the levels of sequence homology between the corresponding ORFs 6, 7, 8 and 9, respectively.

#### Example 2

##### ORF 8 Directs DHA Production

As described in example 1, DNA homologous to *Sp* ORF 6 was found in an unrelated species, SS9 *Photobacter*, which also is capable of producing EPA. Additionally, ORFs homologous to *Sp* ORF 6-9 were found in the DHA producing *Vibrio marinus* (*Vibrio*). From these ORFs a series of experiments was designed in which deletions in each of *Sp* ORFs 6-9 that suppressed EPA synthesis in *E. coli* (Yazawa (1996) *supra*) were complemented by the corresponding homologous genes from *Vibrio*.

The *Sp* EPA cluster was used to determine if any of the *Vibrio* ORFs 6-9 was responsible for the production of DHA. Deletion mutants provided for each of the *Sp* ORFs are EPA and DHA null. Each deletion was then complemented by the corresponding *Vibrio* ORF expressed behind a *lac* promoter (Figure 13).

The complementation of a *Sp* ORF 6 deletion by a *Vibrio* ORF 6 reestablished the production of EPA. Similar results were obtained by complementing the *Sp* ORF 7 and ORF 9 deletions. By contrast, the complementation of a *Sp* ORF 8 deletion resulted in the production of C22:6. *Vibrio* ORF 8 therefore appears to be a key element in the synthesis of DHA. Figures 14 and 15 show chromatograms of fatty acid profiles from the respective complementations of *Sp* del ORF 6 with *Vibrio* ORF 6 (EPA and no DHA) and *Sp* del ORF 8 with *Vibrio* ORF 8 (DHA). Figure 16 shows the fatty acid percentages for the ORF 8 complementation, again demonstrating that ORF 8 is responsible for DHA production.

These data show that polyketide-like synthesis genes with related or similar ORFs can be combined and expressed in a heterologous system and used to produce a distinct PUFA species in the host system, and that ORF 8 has a role in determining the ultimate chain length. The *Vibrio* ORFs 6, 7, 8, and 9 reestablish EPA synthesis. In the case of *Vibrio* ORF 8, DHA is also present (*ca.* 0.7%) along with EPA (*ca.* 0.6%) indicating that this gene plays a significant role in directing synthesis of DHA vs EPA for these systems.

Example 3Requirements for Production of DHA

To determine how *Vibrio* ORFs of the cluster ORF 6-9 are used in combination with *Vibrio* ORF 8, some combinations of *Vibrio* ORF 8 with some or all of the other *Vibrio* ORFS 6-9 cluster were created to explain the synthesis of DHA.

*Vibrio* ORFs 6-9 were complemented with *Sp* ORF 3. The results of this complementation are presented in Figures 16b and 16c. The significant amounts of DHA measured (greater than about 9%) and the absence of EPA suggest that no ORFs other than those of *Vibrio* ORFs 6-9 are required for DHA synthesis when combined with *Sp* ORF 3. This suggests that *Sp* ORF 3 plays a general function in the synthesis of bacterial PUFAs.

With respect to the DHA vs EPA production, it may be necessary to combine *Vibrio* ORF 8 with other *Vibrio* ORFs of the 6-9 cluster in order to specifically produce DHA. The roles of *Vibrio* ORF 9 and each of the combinations of *Vibrio* ORFs (6,8), (7, 8), (8, 9), etc in the synthesis of DHA are being studied.

Example 4Plant Expression Constructs

A cloning vector with very few restriction sites was designed to facilitate the cloning of large fragments and their subsequent manipulation. An adapter was assembled by annealing oligonucleotides with the sequences AAGCCCGGGCTT, SEQ ID NO:44, and GTACAAGCCCGGGCTTAGCT, SEQ ID NO:45. This adapter was ligated to the vector pBluescript II SK+ (Stratagene) after digestion of the vector with the restriction endonucleases *Asp*718 and *Sst*I. The resulting vector, pCGN7769 had a single *Srf*I (and embedded *Sma*I) cloning site for the cloning of blunt ended DNA fragments.

A plasmid containing the napin cassette from pCGN3223, (USPN 5,639,790) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence CGCGATTAAATGGCGCGCCCTGCAGGCGGCCCTGCAGGGCGC GCCATTAAAT, SEQ ID NO:46, was ligated into the vector pBC SK+ (Stratagene) after digestion of the vector with the restriction endonuclease *Bss*HII to construct vector pCGN7765. Plasmids pCGN3223 and pCGN7765 were digested with *Not*I and ligated together. The resultant vector, pCGN7770 (Figure 17), contains the pCGN7765 backbone and the napin seed specific expression cassette from pCGN3223.

*Shewanella* constructs

Genes encoding the *Shewanella* proteins were mutagenized to introduce suitable cloning sites 5' and 3' ORFs using PCR. The template for the PCR reactions was DNA of the cosmid pEPA (Yazawa *et al, supra*). PCR reactions were performed using Pfu DNA polymerase according to the manufacturers' protocols. The PCR products were cloned into *SrfI* digested pCGN7769. The primers CTGCAGCTCGAGACAATGTTGATT  
 5 TCCTTATACTTCTGTCC, SEQ ID NO:47, and GGATCCAGATCTCTAGCTAGTC  
 TTAGCTGAAGCTCGA, SEQ ID NO:48, were used to amplify ORF 3, and to generate plasmid pCGN8520. The primers TCTAGACTCGAGACAATGAGCCAGACCTC  
 TAAACCTACA, SEQ ID NO:49, and CCCGGGCTCGAGCTAATTCGCCTCACTGTC  
 10 GTTTGCT, SEQ ID NO:50, were used to amplify ORF 6, and generate plasmid pCGN7776. The primers GAATTCCTCGAGACAATGCCGCTGCGCATCG  
 CACTTATC, SEQ ID NO: 51, and GGTACCAGATCTTTAGACTTCCCCTTGAAG  
 TAAATGG, SEQ ID NO:52, were used to amplify ORF 7, and generate plasmid pCGN7771. The primers GAATTCGTCGACACAATGTCATTACCAGACAATGC  
 15 TTCT, SEQ ID NO:53, and TCTAGAGTCGACTTATACAGATTCTTCGATGCT  
 GATAG, SEQ ID NO:54, were used to amplify ORF 8, and generate plasmid pCGN7775. The primers GAATTCGTCGACACAATGAATCCTACAGCAACTAACGAA, SEQ ID NO:55, and  
 TCTAGAGGATCCTTAGGCCATTCTTTGGTTTGGCTTC, SEQ ID NO:56, were used to amplify ORF 9, and generate plasmid pCGN7773.

20 The integrity of the PCR products was verified by DNA sequencing of the inserts of pCGN7771, pCGN8520, and pCGN7773. ORF 6 and ORF 8 were quite large in size. In order to avoid sequencing the entire clones, the center portions of the ORFs were replaced with restriction fragments of pEPA. The 6.6 kilobase *PacI/BamHI* fragment of pEPA containing the central portion of ORF 6 was ligated into *PacI/BamHI* digested pCGN7776 to yield  
 25 pCGN7776B4. The 4.4 kilobase *BamHI/BglII* fragment of pEPA containing the central portion of ORF 8 was ligated into *BamHI/BglII* digested pCGN7775 to yield pCGN7775A. The regions flanking the pEPA fragment and the cloning junctions were verified by DNA sequencing.

Plasmid pCGN7771 was cut with *XhoI* and *BglII* and ligated to pCGN7770 after digestion with *SaII* and *BglII*. The resultant napin/ORF 7 gene fusion plasmid was designated  
 30 pCGN7783. Plasmid pCGN8520 was cut with *XhoI* and *BglII* and ligated to pCGN7770 after digestion with *SaII* and *BglII*. The resultant napin/ORF 3 gene fusion plasmid was designated pCGN8528. Plasmid pCGN7773 was cut with *SaII* and *BamHI* and ligated to pCGN7770 after digestion with *SaII* and *BglII*. The resultant napin/ORF 9 gene fusion plasmid was designated pCGN7785. Plasmid pCGN7775A was cut with *SaII* and ligated to pCGN7770 after digestion  
 35 with *SaII*. The resultant napin/ORF 8 gene fusion plasmid was designated pCGN7782. Plasmid pCGN7776B4 was cut with *XhoI* and ligated to pCGN7770 after digestion with *SaII*. The resultant napin/ORF 6 gene fusion plasmid was designated pCGN7786B4.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt (1990) *Plant Molecular Biology*, 14:269-276). The polylinker of pCGN1558 was replaced as a *HindIII*/*Asp718* fragment with a polylinker containing unique restriction endonuclease sites, *AscI*, *PacI*, *XbaI*, *SwaI*, *BamHI*, and *NotI*. The *Asp718* and *HindIII* restriction endonuclease sites are retained in pCGN5139. pCGN5139 was digested with *NotI* and ligated with *NotI* digested pCGN7786B4. The resultant binary vector containing the napin/ORF 6 gene fusion was designated pCGN8533. Plasmid pCGN8533 was digested with *Sse8387I* and ligated with *Sse8387I* digested pCGN7782. The resultant binary vector containing the napin/ORF 6 gene fusion and the napin/ORF 8 gene fusion was designated pCGN8535 (Figure 18).

The plant binary transformation vector, pCGN5139, was digested with *Asp718* and ligated with *Asp718* digested pCGN8528. The resultant binary vector containing the napin/ORF 3 gene fusion was designated pCGN8532. Plasmid pCGN8532 was digested with *NotI* and ligated with *NotI* digested pCGN7783. The resultant binary vector containing the napin/ORF 3 gene fusion and the napin/ORF 7 gene fusion was designated pCGN8534. Plasmid pCGN8534 was digested with *Sse8387I* and ligated with *Sse8387I* digested pCGN7785. The resultant binary vector containing the napin/ORF 3 gene fusion, the napin/ORF 7 gene fusion and the napin/ORF 9 gene fusion was designated pCGN8537 (Figure 19).

#### 20 Vibrio constructs

The *Vibrio* ORFs for plant expression were all obtained using *Vibrio* cosmid #9 as a starting molecule. *Vibrio* cosmid #9 was one of the cosmids isolated from the *Vibrio* cosmid library using the *Vibrio* ORF 6 PCR product described in Example 1.

A gene encoding *Vibrio* ORF 7 (Figure 6) was mutagenized to introduce a *SalI* site upstream of the open reading frame and *BamHI* site downstream of the open reading frame using the PCR primers: TCTAGAGTCGACACAATGGCGGAATTAGCTG TTATTGGT, SEQ ID NO:57, and GTCGACGGATCCCTATTTGTTCTGTTTGCTA TATG, SEQ ID NO:58. A gene encoding *Vibrio* ORF 9 (Figure 6) was mutagenized to introduce a *BamHI* site upstream of the open reading frame and an *XhoI* site downstream of the open reading frame using the PCR primers: GTCGACGGATCCA CAATGAATATAGTAAGTAATCATTCGGCA, SEQ ID NO:59, and GTCGACCTC GAGTTAATCACTCGTACGATAACTTGCC, SEQ ID NO:60. The restriction sites were introduced using PCR, and the integrity of the mutagenized plasmids was verified by DNA sequence. The *Vibrio* ORF 7 gene was cloned as a *SalI*-*BamHI* fragment into the napin cassette of *Sal*-*BglI* digested pCGN7770 (Figure 17) to yield pCGN8539. The *Vibrio* ORF 9 gene was cloned as a *SalI*-*BamHI* fragment into the napin cassette of *Sal*-*BalI* digested pCGN7770 (Figure 17) to yield pCGN8543.

Genes encoding the *Vibrio* ORF 6 and ORF 8 were mutagenized to introduce *Sa*I sites flanking the open reading frames. The *Sa*I sites flanking ORF 6 were introduced using PCR. The primers used were: CCCGGGTCGACACAATGGCTAAAAAGAACA CCACATCGA, SEQ ID NO:61, and CCCGGGTCGACTCATGACATATCGTTCAAA ATGTCCTGA, SEQ ID NO:62. The central 7.3 kb *Bam*HI-*Xho*I fragment of the PCR product was replaced with the corresponding fragment from *Vibrio* cosmid #9. The mutagenized ORF 6 were cloned into the *Sa*I site of the napin cassette of pCGN7770 to yield plasmid pCGN8554.

The mutagenesis of ORF 8 used a different strategy. A *Bam*HI fragment containing ORF 8 was subcloned into plasmid pHC79 to yield cosmid #9". A *Sa*I site upstream of the coding region was introduced on an adapter comprised of the oligonucleotides TCGACATGGAAAATATTGCAGTAGTAGGTATTGCTAATTT GTTC, SEQ ID NO:63, and CCGGGAACAAATTAGCAATACCTACTACTGCAAT ATTTTCCATG, SEQ ID NO:64. The adapter was ligated to cosmid #9" after digestion with *Sa*I and *Xma*I. A *Sa*I site was introduced downstream of the stop codon by using PCR for mutagenesis. A DNA fragment containing the stop codon was generated using cosmid #9" as a template with the primers TCAGATGAACTTTATCGATAC, SEQ ID NO:65 and TCATGAGACGTCGTCGACTTACGCTTCAACAATACT, SEQ ID NO:66. The PCR product was digested with the restriction endonucleases *Cla*I and *Aat*II and was cloned into the cosmid 9" derivative digested with the same enzymes to yield plasmid 8P3. The *Sa*I fragment from 8P3 was cloned into *Sa*I digested pCGN7770 to yield pCGN8515.

pCGN8532, a binary plant transformation vector that contains a *Shewannella* ORF 3 under control of the napin promoter was digested with *Not*I, and a *Not*I fragment of pCGN8539 containing a napin *Vibrio* ORF 7 gene fusion was inserted to yield pCGN8552. Plasmid pCGN8556 (Figure 23), which contains *Shewannella* ORF 3, and *Vibrio* ORFs 7 and 9 under control of the napin promoter was constructed by cloning the *Sse*8357 fragment from pCGN8543 into *Sse*8387 digested pCGN8552.

The *Not*I digested napin/ORF 8 gene from plasmid pCGN8515 was cloned into a *Not*I digested plant binary transformation vector pCGN5139 to yield pCGN8548. The *Sse*8387 digested napin/ORF 6 gene from pCGN8554 was subsequently cloned into the *Sse*8387 site of pCGN8566. The resultant binary vector containing the napin/ORF 6 gene fusion and napin/ORF 8 gene fusion was designated pCGN8560 (Figure 22).

#### Example 5

#### Plant Transformation and PUFA Production

#### EPA production

The *Shewanella* constructs pCGN8535 and pCGN8537 can be transformed into the same or separate plants. If separate plants are used, the transgenic plants can be crossed resulting in heterozygous seed which contains both constructs.

pCGN8535 and pCGN8537 are separately transformed into *Brassica napus*. Plants are  
5 selected on media containing kanamycin and transformation by full length inserts of the constructs is verified by Southern analysis. Immature seeds also can be tested for protein expression of the enzyme encoded by ORFs 3, 6, 7, 8, or 9 using western analysis, in which case, the best expressing pCGN8535 and pCGN8537 T<sub>1</sub> transformed plants are chosen and are grown out for further experimentation and crossing. Alternatively, the T<sub>1</sub> transformed plants  
10 showing insertion by Southern are crossed to one another producing T<sub>2</sub> seed which has both insertions. In this seed, half seeds may be analyzed directly from expression of EPA in the fatty acid fraction. Remaining half-seed of events with the best EPA production are grown out and developed through conventional breeding techniques to provide *Brassica* lines for production of EPA.

15 Plasmids pCGN7792 and pCGN7795 also are simultaneously introduced into *Brassica napus* host cells. A standard transformation protocol is used (*see* for example USPN 5,463,174 and USPN 5,750,871, however *Agrobacteria* containing both plasmids are mixed together and incubated with *Brassica* cotyledons during the cocultivation step. Many of the resultant plants are transformed with both plasmids.

20

#### DHA production

A plant is transformed for production of DHA by introducing pCGN8556 and pCGN8560, either into separate plants or simultaneously into the same plants as described for EPA production.

25 Alternatively, the *Shewanella* ORFs can be used in a concerted fashion with ORFs 6 and 8 of *Vibrio*, such as by transforming with a plant the constructs pCGN8560 and pCGN7795, allowing expression of the corresponding ORFs in a plant cell. This combination provides a PKS-like gene arrangement comprising ORFs 3, 7 and 9 of *Shewanella*, with an ORF 6 derived from *Vibrio* and also an ORF 8 derived from *Vibrio*. As described above, ORF 8 is the PKS-like  
30 gene which controls the identity of the final PUFA product. Thus, the resulting transformed plants produce DHA in plant oil.

#### Example 6

##### Transgenic plants containing the *Shewanella* PUFA genes

35 *Brassica* plants

Fifty-two plants cotransformed with plasmids pCGN8535 and pCGN8537 were analyzed using PCR to determine if the *Shewanella* ORFs were present in the transgenic plants. Forty-one plants contained plasmid pCGN8537, and thirty-five plants contained pCGN8535. 11 of the plants contained all five ORFs required for the synthesis of EPA. Several plants contained genes from both of the binary plasmids but appeared to be missing at least one of the ORFs. Analysis is currently being performed on approximately twenty additional plants.

Twenty-three plants transformed with pCGN8535 alone were analyzed using PCR to determine if the *Shewanella* ORFs were present in the transgenic plants. Thirteen of these plants contained both *Shewanella* ORF 6 and *Shewanella* ORF 8. Six of the plants contained only one ORF.

Nineteen plants transformed with pCGN8537 were alone analyzed using PCR to determine if the *Shewanella* ORFs were present in the transgenic plants. Eighteen of the plants contained *Shewanella* ORF 3, *Shewanella* ORF 7, and *Shewanella* ORF 9. One plant contained *Shewanella* ORFs 3 and 7.

#### Arabidopsis

More than 40 transgenic *Arabidopsis* plants cotransformed with plasmids pCGN8535 and pCGN8537 are growing in our growth chambers. PCR analysis to determine which of the ORFs are present in the plants is currently underway.

#### Example 7

##### Evidence of A PKS System of PUFA Synthesis In *Schizochytrium*

The purpose of this experiment was to identify additional sources of PKS genes. Polyunsaturated long chain fatty acids were identified in *Schizochytrium* oil. Furthermore, production of polyunsaturated fatty acids was detected in a culture of *Schizochytrium*. A freshly diluted culture of *Schizochytrium* was incubated at 24°C in the presence of [<sup>14</sup>C]-acetate (5uCi/mL) for 30 min with shaking (150 rpm). The cells were then collected by centrifugation, lyophilized and subjected to a transesterification protocol that involved heating to 90°C for 90 minutes in the presence of acidic (9% H<sub>2</sub>SO<sub>4</sub>) methanol with toluene (1 volume of toluene per two volumes of acidic methanol) as a second solvent. The resulting methylesters were extracted with an organic solvent (hexane) and separated by TLC (silica gel G, developed three times with hexane:diethyl ether (19:1)). Radioactivity on the TLC plate was detected using a scanner (AMBIS). Two prominent bands were detected on the TLC plate. These bands migrated on the TLC plate in positions expected for short chain (14 to 16 carbon), saturated methyl esters (the upper band) and with methylesters of polyunsaturated long chain (20 to 22 carbon) fatty acids (the lower band). These were also the major types of fatty acids detected by GC analysis of FAMES of *Schizochytrium* oil.



In a parallel experiment thiolactomycin, a well known inhibitor of Type II fatty acid synthesis systems as well as several polyketide synthesis systems including EPA production by *E. coli* transformed with PKS genes derived from *Shewanella*, was added to the test tubes of varying concentrations (0, 1, 10 and 100 µg/ml) prior to addition of the *Schizochytrium* cell cultures and [<sup>14</sup>C] acetate. Analysis of incorporation of [<sup>14</sup>C] acetate, as described above, revealed that 100 ug/mL thiolactomycin completely blocked synthesis of polyunsaturated fatty acids, while partial inhibition of synthesis of polyunsaturated fatty acids was observed at 10 ug/mL thiolactomycin. Synthesis of the short chain saturated fatty acids was unaffected at all tested thiolactomycin concentrations. Thiolactomycin does not inhibit Type I fatty acid synthesis systems and is not toxic to mice, suggesting that it does not inhibit the elongation system leading to EPA or DHA formation. Furthermore, thiolactomycin did not inhibit the elongation system leading to PUFA synthesis in *Phaeodactylum tricornutum*. Therefore, although *Schizochytrium* is known to possess a Type I fatty acid synthesis system, the data suggested that the polyunsaturated fatty acids produced in this organism were derived from a system which was distinct from the Type I fatty acid synthesis system which produced short chain fatty acids, and from a system that was similar to the elongation/desaturation pathway found in mice and *Phaeodactylum*. The data are consistent with DHA formation being a result of a PKS pathway as found in *Vibrio marinus* and *Shewanella putrefaciens*.

#### Example 8

##### PKS Related Sequences From *Schizochytrium*

The purpose of this experiment was to identify sequences from *Schizochytrium* that encoded PKS genes. A cDNA library from *Schizochytrium* was constructed and approximately 8,000 random clones (ESTs) were sequenced. The protein sequence encoded by *Shewanella* EPA synthesis genes was compared to the predicted amino acid sequences of the *Schizochytrium* ESTs using a Smith/Waterman alignment algorithm. When the protein sequence of ORF6 (*Shewanella*) was compared with the amino acid sequences from *Schizochytrium* ESTs, 38 EST clones showed a significant degree of identity ( $P < 0.01$ ). When the protein sequence of ORF7 was compared by *Schizochytrium* ESTs, 4 EST clones showed significant identity ( $P < 0.01$ ) suggesting that the molecules were homologous. When the protein sequence of ORF8 and ORF9 were compared with the *Schizochytrium* ESTs, 7 and 14 clones respectively showed significant identity ( $P < 0.01$ ).

#### Example 9

##### Analysis of *Schizochytrium* cDNA Clones

Restriction enzyme analysis of the *Schizochytrium* EST clones was used to determine the longest clones, which were subsequently sequenced in their entirety. All of the EST sequences described in Example 8 were determined to be part of 5 cDNA clones.

Two of the cDNA clones were homologous to *Shewanella* ORF6. LIB3033-047-B5 was  
5 homologous to the C-terminus of ORF6. The sequence of LIB3033-047-B5 could be aligned with *Shewanella* ORF6 from amino acids 2093 onwards. The open reading frame of LIB3033-047-B5 extended all the way to the 5' end of the sequence, thus this clone was not likely to be full length. LIB3033-046-E6 shared homology to the ACP domain of ORF6. It contained 6  
10 ACP repeats. This cDNA clone did not have a poly-A-tail, and therefore, it was likely to be a partial cDNA with additional regions of the cDNA found downstream of the sequence. The PCR primers GTGATGATCTTTCCCTGATGCACGCCAAGG (SEQ ID NO: 67) and AGCTCGAGACCGGCAACCCGCAGCGCCAGA (SEQ ID NO: 68) were used to amplify a fragment of approximately 500 nucleotides from *Schizochytrium* genomic DNA. Primer GTGATGATCTTTCCCTGATGCACGCCAAGG was derived from LIB3033-046-E6, and  
15 primer AGCTCGAGACCGGCAACCCGCAGCGCCAGA was derived from LIB3033-047-B5. Thus, LIB3033-046-E6 and LIB3033-047-B5 represented different portions of the same mRNA (see Figure 28) and could be assembled into a single partial cDNA sequence (see Figure 27A), SEQ ID NO: 69, that was predicted to encode a protein with the sequence in Figure 29A (SEQ ID NO: 70). The open reading frame extended all the way to the 5' end of the sequence, thus this  
20 partial cDNA was not likely to be full length. Analysis of additional cDNA or genomic clones will allow the determination of the full extent of the mRNA represented by clones LIB3033-046-E6 and LIB3033-047-B5. It may contain condensing enzyme related domains similar to those found near the N-terminus of *Shewanella* ORF6.

One of the cDNA clones, LIB3033-046-D2, was homologous to *Shewanella* ORF9 at its  
25 3' end. This clone was homologous to the chain length factor region of *Shewanella* ORF8 at its 5' end. This clone was also homologous to the entire open reading frame of the *Anabaena* HglC ORF. The *Anabaena* HglC ORF is homologous to the chain length factor region of *Shewanella* ORF8 and *Shewanella* ORF7. Thus this cDNA (Figure 27B), SEQ ID NO: 71, was homologous to part of *Shewanella* ORF8, *Shewanella* ORF7 and *Shewanella* ORF9 (see Figure 28). The  
30 amino acid sequence (Figure 29B), SEQ ID NO: 72, encoded by the open reading frame of LIB3033-046-D2 extended all the way to the 5' end of the sequence; thus this clone was not likely to be full length. Analysis of additional cDNA or genomic clones will allow the determination of the full extent of the mRNA represented by LIB3033-046-E6. It may contain condensing enzyme related domains similar to those found near the N-terminus of *Shewanella*  
35 ORF8.

Two additional cDNA clones were homologous to *Shewanella* ORF8. LIB81-015-D5 was homologous to the C-terminus of ORF8. The 5' sequence of LIB81-015-D5 could be

aligned with *Shewanella* ORF8 from amino acids 1900 onwards. The 3' end of LIB81-015-D5 could be aligned with *Shewanella* ORF9 (see Figure 28). The amino acid sequence (Figure 29C), SEQ ID NO: 73, encoded by the open reading frame of LIB81-015-D5 extended all the way to the 5' end of the sequence; thus this clone was not likely to be full length. LIB81-042-B9 was homologous to amino acids 1150 to 1850 of *Shewanella* ORF8. LIB81-042-B9 did not have a poly-A-tail, and therefore, it was likely to be a partial cDNA with additional regions of the cDNA found downstream of the sequence. The PCR primers TACCGCGGCAAGACTATCCGCAACGTCACC (SEQ ID NO: 74) and GCCGTCGTGGGCGTCCACGGACACGATGTG (SEQ ID NO: 75) were used to amplify a fragment of approximately 500 nucleotides from *Schizochytrium* genomic DNA. Primer TACCGCGGCAAGACTATCCGCAACGTCACC was derived from LIB81-042-B9, and primer GCCGTCGTGGGCGTCCACGGACACGATGTG was derived from LIB81-015-D5. Thus, LIB81-042-and LIB81-015-D5 represented different portions of the same mRNA and were assembled into a single partial cDNA sequence (see Figure 27C), SEQ ID NO: 76. The open reading frame of LIB81-042-B9 also extended all the way to the 5' end of the sequence, thus this clone was also not likely to be full length. Analysis of additional cDNA or genomic clones will allow the determination of the full extent of the mRNA represented by LIB81-042-B9.

By the present invention PKS-like genes from various organisms can now be used to transform plant cells and modify the fatty acid compositions of plant cell membranes or plant seed oils through the biosynthesis of PUFAs in the transformed plant cells. Due to the nature of the PKS-like systems, fatty acid end-products produced in the plant cells can be selected or designed to contain a number of specific chemical structures. For example, the fatty acids can comprise the following variants: Variations in the numbers of keto or hydroxyl groups at various positions along the carbon chain; variations in the numbers and types (*cis* or *trans*) of double bonds; variations in the numbers and types of branches off of the linear carbon chain (methyl, ethyl, or longer branched moieties); and variations in saturated carbons. In addition, the particular length of the end-product fatty acid can be controlled by the particular PKS-like genes utilized.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

What is claimed is:

1. An isolated nucleic acid comprising:  
a *Vibrio marinus* nucleotide sequence selected from the group consisting of ORF 6 (SEQ ID NO:77), ORF 7 (SEQ ID NO:78), ORF 8 (SEQ ID NO:79), and ORF 9 (SEQ ID NO:80), as  
5 shown in Figure 6.
2. An isolated nucleic acid comprising:  
a nucleotide sequence which encodes a polypeptide of a polyketide-like synthesis  
system, wherein said system produces a docosahexenoic acid when expressed in a host cell.  
10
3. The isolated nucleic acid according to Claim 2, wherein said nucleotide sequence is  
derived from a marine bacterium.
4. An isolated nucleic acid according to Claim 2, wherein said nucleotide sequence is  
15 derived from *Schizochytrium*.
5. The isolated nucleic acid according to Claim 2, wherein said nucleotide sequence is a  
*Vibrio marinus* ORF 8 (SEQ ID NO:79), as shown in Figure 6.
- 20 6. An isolated nucleic acid comprising a *Schizochytrium* nucleotide sequence comprising a  
sequence shown in a SEQ ID NO selected from the group consisting of SEQ ID NOS: 69, 71  
and 76.
7. An isolated nucleic acid comprising:  
25 a nucleotide sequence which is substantially identical to a sequence of at least 50  
nucleotides of a *Vibrio marinus* nucleotide sequence selected from the group consisting of ORF  
6 (SEQ ID NO:77), ORF 7 (SEQ ID NO:78), ORF 8 (SEQ ID NO:79), and ORF 9 (SEQ ID  
NO:80), as shown in Figure 6.
- 30 8. A recombinant microbial cell comprising at least one copy of an isolated nucleic acid  
according to Claim 6.
9. The recombinant microbial cell according to Claim 8, wherein said cell comprises each  
element of a polyketide-like synthesis system required to produce a long chain polyunsaturated  
35 fatty acid.

10. The recombinant microbial cell according to Claim 9, wherein said cell is a eukaryotic cell.
11. The recombinant microbial cell according to Claim 10, wherein said eukaryotic cell is a  
5 fungal cell, an algae cell or an animal cell.
12. The recombinant microbial cell according to Claim 11, wherein said fungal cell is a yeast cell and said algae cell is a marine algae cell.
- 10 13. The recombinant microbial cell according to Claim 8, wherein said cell is a prokaryotic cell.
14. The recombinant microbial cell according to Claim 13, wherein said cell is a bacterial cell or a cyanobacterial cell.
- 15 15. A recombinant cell according to Claim 14, wherein said bacterial cell is a *lactobacillus* cell.
16. The microbial cell according to Claim 8, wherein said recombinant microbial cell is  
20 enriched for 22:6 fatty acids as compared to a non-recombinant microbial cell which is devoid of said isolated nucleic acid.
17. A method for production of docosahexenoic acid in a microbial cell culture, said method comprising:  
25 growing a microbial cell culture having a plurality of microbial cells, wherein said microbial cells or ancestors of said microbial cells were transformed with a vector comprising one or more nucleic acids having a nucleotide sequence which encodes a polypeptide of a polyketide synthesizing system, wherein said one or more nucleic acids are operably linked to a promoter, under conditions whereby said one or more nucleic acids are expressed and  
30 docosahexenoic acid is produced in said microbial cell culture.
18. A method for production of a long chain polyunsaturated fatty acid in a plant cell, said method comprising:  
growing a plant having a plurality of plant cells, wherein said plant cells or ancestors of  
35 said plant cells were transformed with a vector comprising one or more nucleic acids having a nucleotide sequence which encodes one or more polypeptides of a polyketide synthesizing system which produces a long chain polyunsaturated fatty acid, wherein each of said nucleic

acids are operably linked to a promoter functional in a plant cell, under conditions whereby said polypeptides are expressed and a long chain polyunsaturated fatty acid is produced in said plant cells.

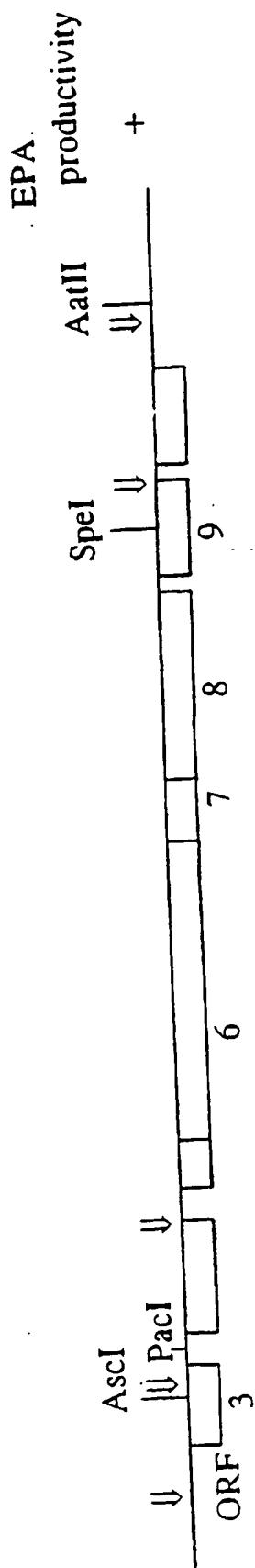
- 5 19. The method according to Claim 17 or Claim 18 wherein said nucleotide sequence is shown in a SEQ ID NO selected from the group consisting of SEQ ID NOS: 69, 71 and 76.
20. The method according to Claim 18, wherein said long chain polyunsaturated fatty acid produced in said plant cells is a 20:5 and 22:6 fatty acid.
- 10 21. The method according to Claim 17, wherein said nucleotide sequence is selected from the group consisting of *Vibrio marinus* ORF 6 (SEQ ID NO:77), ORF 7 (SEQ ID NO:78), ORF 8 (SEQ ID NO:79), and ORF 9 (SEQ ID NO:80), as shown in Figure 6 and *Shewanella putrefaciens* ORF 6 (SEQ ID NO:83), ORF 7 (SEQ ID NO:84), ORF 8 (SEQ ID NO:85), ORF 9  
15 (SEQ ID NO:86), and ORF 3, which is complementary to SEQ ID NO:4, as shown in Figure 4.
22. The method according to Claim 18, wherein said nucleic acid constructs are derived from two or more polyketide synthesizing systems.
- 20 23. The method according to Claim 18, wherein said long chain polyunsaturated fatty acid is eicosapentenoic acid.
24. The method according to Claim 18, wherein said long chain polyunsaturated fatty acid is docosahexenoic acid.
- 25 25. A recombinant plant cell comprising:  
one or more nucleic acids having a nucleotide sequence which encodes one or more polypeptides of a polyketide synthesizing system which produces a long chain polyunsaturated fatty acid, wherein each of said nucleic acids are operably linked to a promoter functional in said  
30 plant cell.
26. The recombinant plant cell according to Claim 25, wherein said nucleotide sequence is shown in a SEQ ID NO selected from the group consisting of SEQ ID NOS: 69, 71 and 76.
- 35 27. The recombinant plant cell according to Claim 26, wherein said recombinant plant cell is a recombinant seed cell.

28. The recombinant plant cell according to Claim 27, wherein said recombinant seed cell is a recombinant embryo cell.
29. The recombinant plant cell according to Claim 26, wherein said recombinant plant cell is  
5 from a plant selected from the group consisting of *Brassica*, soybean, safflower, and sunflower.
30. A plant oil produced by a recombinant plant cell according to Claim 26.
31. The plant oil according to Claim 30, wherein said plant oil comprises eicosapentenoic  
10 acid.
32. The plant oil according to Claim 30, wherein said plant oil comprises docosahexenoic acid.
- 15 33. The plant oil according to Claim 30, wherein said plant oil is encapsulated.
34. A dietary supplement comprising a plant oil according to Claim 30.
35. A recombinant *E. coli* cell comprising:  
20 one or more nucleic acids having a nucleotide sequence which encodes one or more polypeptides of a polyketide synthesizing system which produces a long chain polyunsaturated fatty acid, wherein each of said nucleic acids are operably linked to a promoter function in said *E. coli* cell.
- 25 36. The recombinant *E. coli* cell according to Claim 35, wherein said long chain polyunsaturated fatty acid is docosahexenoic acid.
37. The recombinant *E. coli* cell according to Claim 35, wherein said nucleotide sequence is shown in a SEQ ID NO selected from the group consisting of SEQ ID NOS: 69, 71 and 76.  
30
38. A plant oil produced by a recombinant plant cell wherein said plant oil comprises a long chain polyunsaturated fatty acid exogenous to said plant oil, wherein said plant cell is produced according to a method comprising:  
transforming said plant cell or an ancestor of said plant cell with a vector comprising one  
35 or more polypeptide of a polyketide synthesizing system which produces a long chain polyunsaturated fatty acid wherein each of said nucleic acids are operably linked to a promoter functional in said plant cell.



39. A plant oil according to Claim 38, wherein said long chain polyunsaturated fatty acid is eicosapentenoic acid.

- 5 40. A plant oil according to Claim 38, wherein said long chain polyunsaturated fatty acid is docosahexenoic acid.



1/134

FIG. 1A

2/134

EPA  
productivity

+

pAA-NEB

AscI-AatII/NEB

pPA-NEB ( $\Delta 2,3$ )

PacI-AatII/NEB

Single ORF clones

ORF3 / pSTV 28

ORF 6 / pUC118

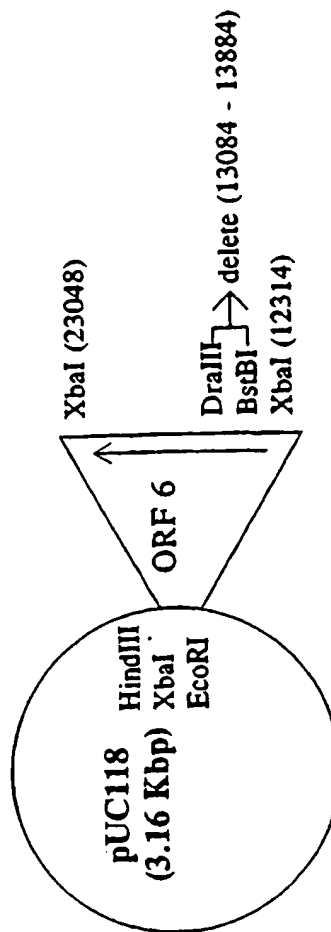
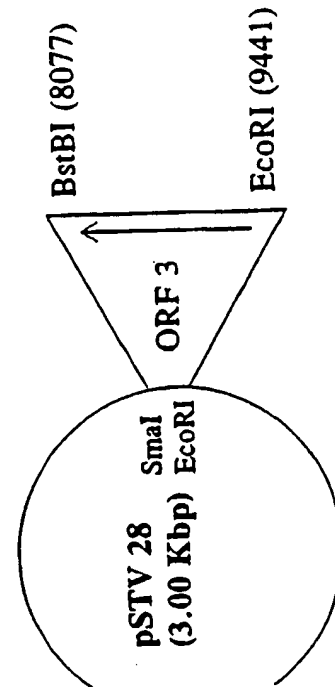


FIG. 1B

Orf6 8.3 KB - 293 kD

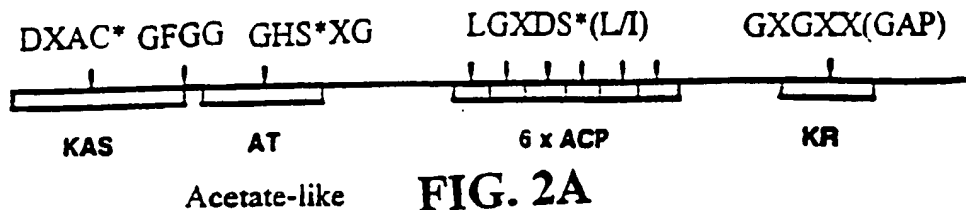


FIG. 2A

Orf7 2.3 KB - 84 kD

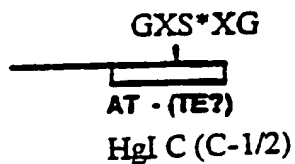


FIG. 2B

Orf3 0.8 KB - 30 kD



FIG. 2E

Orf8 6.0 KB - 217 kD

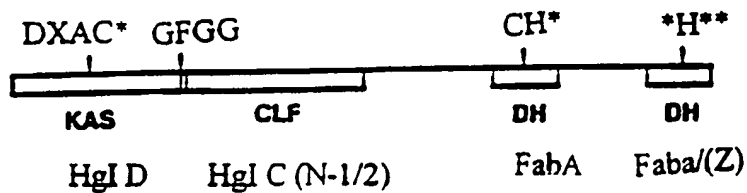


FIG. 2C

Orf9 1.6 KB - 59 kD

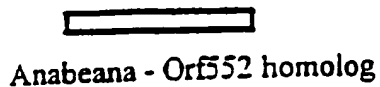
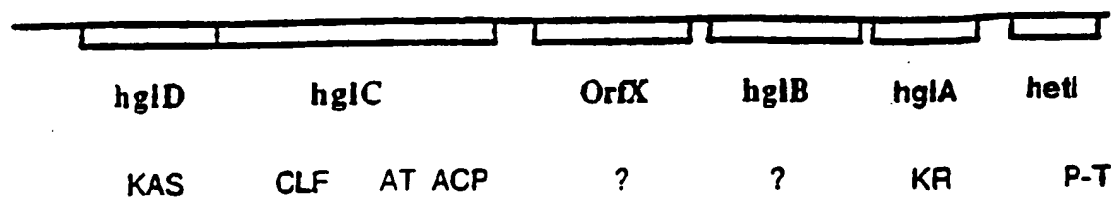
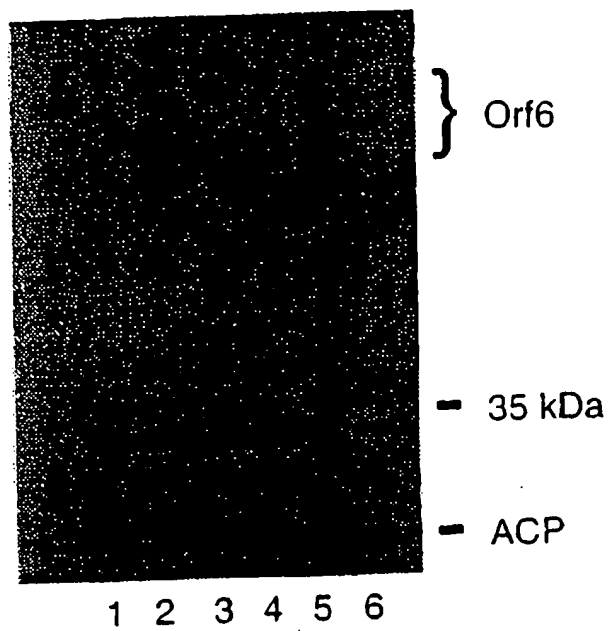


FIG. 2D

4/134

**FIG. 2F**

5/134



**FIG. 3**

6/134

GATCTCTTAC AAAGAAACTA TCTCAATGTG AATTTAACCT TAATTCGGTT TAATTACGGC 60  
CTGATAGAGC ATCACCCCAAT CAGCCATAAA ACTGTAAAGT GGTACTCAA AGGTGGCTGG 120  
GCGATTCTTC TCAAAATACAA AGTGCCCAAC CCAAGCAAAT CCATATCCGA TAACAGGTAA 180  
AAGTAGCAAT AAACCCCGAGC GCTGAGTTAG TAATACATAA GCGAATAATA GGATCACTAA 240  
ACTACTGCCG AAATAGTGTA ATATTCGACA GTTCTATGC TGATGTTGAG ATAAATAAAA 300  
AGGGTAAAAAT TCAGCAAAAAG AACGATAGCG CTTACTCAT TACTCACAACCT CGGTAAAAAA 360  
GCAACTCGCC ATTAACTTGG CCAATCGTCA GTTGTCTCTAT CGTCTCAAAG TTATGCCGAC 420  
TAAATAAATC TATATGTGCA TTATGATTAG CAAAACTCC GATACCATCA AGATGAAAGT 480  
GTTTCATCACA CCAACTCAAA ACTGCGTCGA TAAGCTTACT GCCATAGCCC TTGCCCTTGCT 540  
CCACATTGTC GATAGCAATA AACTGTAAAA TGCCACATTG GCCACTTGGT AAGCTCTCTA 600  
TAATCTGATT TTCCTTTGTTA ATAAGTGCCT GAGTTGAATA CCAACCAGTA CTTAACAACA 660  
TCCTTTAAACG CCAATGCCAA AAACGCGCTT CACCTAAGGG AACCTGCTGA GTCACATATGC 720  
AGGCTACGCC TATCAATCTA TCCCAACGA ACATACCAAT AAGTGCTTGC TCCTGTTGCC 780  
AGAGCTCATT GAGTTCCTTCT CGAATAGCCC CGCGAAGCTT TTGCTCATAC TGGCCTTGAT 840  
CACCACATAA AAGTGTTTCG ATAAAAAAGG GATCATCATG ATAGGCGTTA TAGAGAATAG 900  
AGGCTGCTAT GCGTAAATCT TCTGCCGCTGA GATAAACTGC ACGACACTCT TCCATGGCTT 960  
GATCTTCCAT TGTATTGTC CTTGACCTTG ATCACACAAC ACCAATGTAA CAAGACTGTA 1020

FIG. 4A-1

7/134

TAGAAGTGCA ATTAATAATC AATTCGTGCA TTAAGCAGGT CAGCATTTCT TTGCTAAACA 1080  
AGCTTTATTG GCTTTGACAA AACTTTGCCT AGACTTTAAC GATAGAAATC ATAATGAAAG 1140  
AGAAAAGCTA CAACCTAGAG GGGAAATAATC AAACAACCTGC TAAGATCTAG ATAATGTAAT 1200  
AAACACCGAG TTTATCGACC ATACTTAGAT AGAGTCATAG CAACGAGAAT AGTTATGGAT 1260  
ACAAAGCCGC AAGATCTATC ACACCTGTTT TTACAGCTAG GATTAGCAA TGATCAACCC 1320  
GCAATTGAAC AGTTTATCAA TGACCATCAA TTAGCGGACA ATATATTGCT ACATCAAGCA 1380  
AGCTTTTGGA GCCCATCGCA AAAGCACTTC TTAATTGAGT CATTTAATGA AGATGCCCCAG 1440  
TGGACCGAAG TCATCGACCA CTTAGACACC TTATTAAGAA AAAACTAACC ATTACAACAG 1500  
CAACTTTAAA TTTTGCCGTA AGCCATCTCC CCCACCCCCA CAACAGCGTT GTTGCTTTATG 1560  
ACCACTGGAG TACATTGTC TTTAGTCGTT TTACCATCAC CATGGGTACG TTGAGTGCGA 1620  
TAAAAAAGCA CATAAACTTC TTTATCGGCC TGAATATAGG CTTCGTTAAA ATCAGCTGTT 1680  
CCCATTAAAG TAACCACTTG CTCCTTACTC ATGCCCTAGAG ATATCTTTGT CAAATTGTCA 1740  
CGGTTTTTAT CTTGAGTTTT CTCCCAAGCA CCGTGATTAT CCCAGTCAGA TTCCCCCATCA 1800  
CCAACATTGA CCACACAGCC CGTTAGCCCT AAGCTTGCAA TCCCAAAACA TGCTAAACCT 1860  
AATAAATTTAT TTTTCATTTT AACTTCCTGT TATGACATTA TTTTTCCTTA GAAGAAAAGC 1920  
AACTTACATG CCAAAACACA AGCTGTTGTT TTAATGACT TTATTTATTA TTAGCCTTTT 1980  
AGGATATGCC TAGAGCAATA ATAATTACCA ATGTTTAAGG AATTGACTA ACTATGAGTC 2040

FIG. 4A-2



8/134

CGATTGAGCA AGTGCTAACA GCTGCTAAAA AAATCAATGA ACAAGGTAGA GAACCAACAT 2100  
TAGCATTGAT TAAAAACCAA CTTGGTAATA GCATCCCAAT GCGCGAGTTA ATCCAAGGTT 2160  
TGCAACAGTT TAAGTCTATG AGTGCAGAAG AAAGACAAGC AATACCTAGC AGCTTAGCAA 2220  
CAGCAAAAGA AACTCAATAT GGTCAATCAA GCTTATCTCA ATCTGAACAA GCTGATAGGA 2280  
TCCTCCAGCT AGAAAAACGCC CTCAATGAAT TAAGAAAACGA ATTTAATGGG CTAAAAAGTC 2340  
AATTGATAA CTTACAACAA AACCTGATGA ATAAAGAGCC TGACACCCAAA TGCATGTAAT 2400  
TGAACACTACGA TTTGAATGTT TTGATAACAC CACGATTACT GCAGCAGAAA AAGCCATTAA 2460  
TGGTTTGCTT GAAGCTTATC GAGCCAATGG CCAGGTTCTA GGTGCTGAAT TTGCCCGTTGC 2520  
ATTTAACGAT GGTGAGTTTA AAGCACGCAT GTTAACCCCA GAAAAAAGCA GCTTATCTAA 2580  
ACGCTTTAAT AGTCCTTGGG TAAATAGTGC ACTCGAAGAG CTAACCGAAG CCAAATTGCT 2640  
TGCGCCACGT GAAAAGTATA TTGGCCAAGA TATTAATTCT GAAGCATCTA GCCAAGACAC 2700  
ACCAAGTTGG CAGCTACTTT ACACAAAGTTA TGTGCACATG TGCTCACCAC TAAGAAATGG 2760  
CGACACCTTG CAGCCTATTC CACTGTATCA AATTCCAGCA ACTGCCAACG GCGATCATAA 2820  
ACGAATGATC CGTTGGCAAA CAGAAATGGCA AGCTTGTGAT GAATTGCAAA TGGCCGCAGC 2880  
TACTAAAGCT GAATTTGCCG CACTTGAAGA GCTAACCAGT CATCAGAGTG ATCTATTAG 2940  
GCGTGGTTGG GACTTACGTG GCAGAGTCGA ATACTTGACG AAAATTCCGA CCTATTACTA 3000  
TTTATACCGT GTTGGCGGTG AAAGCTTAGC AGTAGAAAAG CAGCGCTCTT GTCCTAAGTG 3060

FIG. 4A-3

9/134

TGGCAGTCAA GAATGGCTGC TCGATAAACC ATTATTGGAT ATGTTCCATT TTCGCTGTGA 3120  
CACCTGCCGC ATCGTATCTA ATATCTCTTG GGACCATTTA TAACTCTTCC GAGTCTTATC 3180  
ACACTAGAGT TTAGTCAGCA TAAAAATGGC GCTTATATTT CAATTAAAAG AAATATAAGC 3240  
GCCATTTTCA TCGATACTAT ATATCAGCAG ACTATTTTCC GCGTAAATTA GCCCACATTA 3300  
ATTTCAATTCT TTGCCAGATC CCTGGATGAT CTAGTTGTGG CATCGACTCT TCAATAGGTT 3360  
TAACCCGAGG TGTAACCCCTT GGAGTCAATT CGTTTATAAA CTCGTTTAAA CTGTCACTTA 3420  
ATTTAACGCT TTGTACTTCA CCTGGAATTT CAATCCATAC GTGCCATCA CTATTATTAA 3480  
CCGTCAACAT TTTATCTTCA TCATCAAGAA TACCAATAAA CCAAGTCGGC TCTTGCTTAA 3540  
GCTTTCCTT CATCATTTAA TGACCAATGA TGTTTGTG TAAGTATTCA AAATCAGTTT 3600  
GATCCACAC TTGGATTAGC TCACCTTGGC CCCATTGTGA GTCAAAAAAT AGCGGTGCAG 3660  
AAAAATGACT GCCAAAAAAT GGATTAATTT CTGCAGATAA TGTCATTTCAGT 3720  
CAACATTAGC AAATTCACCA GGTGTGTGAC GTACAACCGA TTGCCAAAAC ACTGCGCCAT 3780  
CGGAGCCCGC TTCGGCGACA ACACACTCAG ACTTTGTCC TTGCGCATAA TATCTTGGCT 3840  
GTTACCCAAG CTTATCCATG TAGGCTTGTG GATATTTAGA TAAAAAAGA TCTAAAGCAG 3900  
GTAAAGAAGA CACTTAAGCC AGTTCCAAA TCAGTTATAA TAGGGGTCTA TTTTGACATG 3960  
GAAACCGTAT TGATGACACA ACATCATGAT CCCTACAGTA ACGCCCCCGA ACTTCTGAA 4020  
TTAACTTTAG GAAAGTCGAC CGGTTATCAA GAGCAGTATG ATGCATCTTT ACTACAAGCG 4080

FIG. 4A-4

10/134

TGCCGCGTAA ATTAAACCGT GATGCTATCG GTCTAACCAC TGAGCTACCT TTTCATGGCT 4140  
GTGATATTTG GACTGGCTAC GAACTGTCTT GGCTAAATGC TAAAGGCAAG CCAATGATTG' 4200  
CTATTGCAGA CTTTAACCTA AGTTTGTATA GTAAAAATCT GATCGAGTCT AAGTCGTTTA 4260  
AGCTGTATTT AAACAGCTAT AACCAAACAC GATTTGATAG CGTTCAAGCG GTTCAAGAAC 4320  
GTTTAACTGA AGACTTAAGC GCCTGTGCCC AAGGCACAGT TACGGTAAAA GTGATTGAAC 4380  
CTAAGCAATT TAACCACCTG AGAGTGGTTG ATATGCCAGG TACCTGCATT GACGATTAG 4440  
ATATTGAAGT TGATGACTAT AGCTTTAACT CTGACTATCT CACCGACAGT GTTGATGACA 4500  
AAGTCATGGT TGCTGAAACG CTAACGTCAA ACTTATTGAA ATCAAACCTGC CTAATCACTT 4560  
CTCAGCCTGA CTGGGGTACA GTGATGATCC GTTATCAAGG GCCTAAGATA GACCGTGAAA 4620  
AGCTACTTAG ATATCTGATT TCATTTAGAC AGCACAAATGA ATTTCATGAG CAGTGTGTTG 4680  
AGCGTATATT TGTGTATTTA AAGCACTATT GCCAATGTGC CAAACTTACT GTCTATGCAC 4740  
GTTATACCCG CCGTGGTGGT TTAGATATCA ACCCATATCG TAGCGACTTT GAAAACCCCTG 4800  
CAGAAAAATCA GCGCCTAGCG AGACAGTAAT TGATTGCAGT ACCTACAAA AACCAATGCCT 4860  
ATAAGCCCAAG CTTATGGGCA TTTTATATTT ATCAACTTGT CATCAAACTT CAGCCGCCAA 4920  
GCCTTTTAGT TTTATCGCTA AATTAAAGCCG CTCTCTCAGC CAAATATTTG CAGGATTTTG 4980  
CTGTAATTTA TGGCTCCACA CCATGAAATA CTCTATCGGC TCTACCGCAA AAGGTAAGTC 5040  
AAATACCTGT AAGCCAAAACA GCTTGGCATA TTCGTCAAGT TGGGCTTTTG ACGCGATAGC 5100

FIG. 4A-5

11/134

TAACGCATCA CTTTTTGAGG CAACCGACAT CATACTTAAT ATTGATGATT GCTCGCTGTG 5160  
CATTTGCCCTT GCCGGTAACA CCTGTTTAGT CAGCAAGTCG GCAACACTTA AATTGTAGCG 5220  
GCGCATCTTA AAAATAATAT GCTTTTCATT AAAGTATTGC TCTTGGCTCA ACCCACCTTG 5280  
GATCCTTGGG TGAGCATTTT GTGCCACACA AACTAATTTA TCCTGCATTA CTTTTTGACT 5340  
CTTAAATGCC GCAGATTCTG GCAGCCAAAT ATCTAAGGCT AAATCCACCT TTTCTAGTTG 5400  
TAGGTCCATC TGCAACTCTT CTTCAATGAG CGGCGGCTCA CGAAATACAA TATTAATTGC 5460  
AGTGCCCTGT AACACTTGCT CAATTGATC TTGCAAGAGT TGTATTGCCG ACTCGCTGGC 5520  
ATACACATAA AAAGTTCGCT CACTTGAAGT GGGGTCAAAT GCTTCAAAAGC TAGTCGCAAC 5580  
TTGCTCAATT GTTGACATAG CGCCCGCGAG CTGTTGATAA AGCGTCATCG CACTTGCGGT 5640  
AGGTTTAACT CCCCTACCCA CTCGAGTAAA CAACTCTTCT CCAACAATAC TTTTGTAGCCT 5700  
CGAAATCGCA TTACTAACCG ACGACTGAGT CAAATCCAGC TCTTCTGCCG CCCGGCTAAA 5760  
AGATGAGGTG CGATACACCG CAGTAAAAAC GCGAAATAAA TTAAGATCAA AAGCTTTTG 5820  
CTGCGACATA AATCAGCTAT CTCCTTATCC TTATCCTTAT CCTTATAAAA AGTTAGCTCC 5880  
AGAGCACTCT AGCTCAAAA CAACTCAGCG TATTAAGCCA ATATTTTGGG AACTCAATTA 5940  
ATATTCAATA TAAAAGTATT CATAATATAA ATACCAAGTC ATAATTTAGC CCTAATTATT 6000  
AATCAATTCA AGTTACCTAT ACTGGCCTCA ATTAAGCAA TGTCTCATCA GTCTCCCTGC 6060  
AACTAAATGC AATATTGAGA CATAAAGCTT TGAACGTGATT CAATCTTACG AGGGTAACCT 6120

FIG. 4A-6

12/134

ATGAAACAGA CTCTAATGGC TATCTCAATC ATGTGGCTTT TTTCAATCAA TGCGCTAGCA 6180  
GCGCAACATG AACATGACCA CATCACTGTT GATTACGAAG GGAAAGCCGC AACAGAACAC 6240  
ACCATAGCTC ACAACCAAGC TGTAGCTAAA ACACCTTAACT TTGCCGACAC GCGTGCAATT 6300  
GAGCAATCGT CTAATAATCT AGTCGCCAAG TTGATAAAG CAACTGCCGA TATATTACGT 6360  
GCCGAATTG CTTTATTAG CGATGAAATC CCTGACTCGG TTAACCCGTC TCTCTACCGT 6420  
CAGGCTCAGC TTAATATGGT GCCTAATGGT CTGTATAAAG TGAGCGATGG CATTTACCAG 6480  
GTCCGCGGTA CCGACTTATC TAACCTTACA CTTATCCGCA GTGATAACGG TTGGATAGCA 6540  
TAGGATGTTT TGTAAACCA AGAAGCAGCA AAAGCCCTCAC TACAATTGTC GTTAAAGAAT 6600  
CTACCTAAAG ATGGCGATTT ACCCGTTGTT GCGATGATTT ACTCCCATAG CCATGCGGAC 6660  
CACTTTGGCG GAGCTCGCGG TGTTCAAGAG ATGTTCCCTG ATGTCAAAGT CTACGGCTCA 6720  
GATAACATCA CTAAAGAAAT TGTCGATGAG AACGTACTTG CCGGTAACGC CATGAGCCGC 6780  
CGCGCAGCTT ATCAATACGG CGCAACACTG GGCAACACATG ACCACGGTAT TGTGATGCT 6840  
GCGCTAGGTA AAGGTCTATC AAAAGGTGAA ATCACTTACG TCGCCCCAGA CTACACCTTA 6900  
AACAGTGAAG GCAAAATGGA AACGCTGACG ATTGATGGTC TAGAGATGGT GTTTATGGAT 6960  
GCCTCGGGCA CCGAAGCTGA GTCAGAAATG ATCACTTATA TTCCCTCTAA AAAAGCGCTC 7020  
TGGACGGCGG AGCTTACCTA TCAAGGTATG CACAACATTT ATACGCTGCG CGGCGCTAAA 7080  
GTACGTGATG CGCTCAAGTG GTCAAAAGAT ATCAACGAAA TGATCAATGC CTTTGGTCAA 7140

FIG. 4A-7

13/134

GATGTCGAAG TGCTGTTTGC CTCGCACTCT GCGCCAGTGT GGGGTAACCA AGCGATCAAC 7200  
GATTTCTTAC GCCTACAGCG TGATAACTAC GGCCTAGTGC ACAATCAAAC CTTGAGACTT 7260  
GCCAACGATG GTGTGGGTAT ACAAGATATT GCGGATGCGA TTCAAGACAC GATTCCAGAG 7320  
TCTATCTACA AGACGTGGCA TACCAATGGT TACCACGGCA CTTATAGCCA TAACGCTAAA 7380  
GCGGTTTATA ACAAATATCT AGGCTACTTC GATATGAACC CAGCCAACCT TAATCCGCTG 7440  
CCAACCAAGC AAGAATCTGC CAAGTTTGTG GAATACATGG GCGGCGCAGA TGCCGCAATT 7500  
AAGCGCGCTA AAGATGATTA CGCTCAAGGT GAATACCGCT TTGTTGCAAC GGCATTAAAT 7560  
AAGGTGGTGA TGGCCGAGCC AGAAATGAC TCCGCTCGTC AATTGCTAGC CGATACCTAT 7620  
GAGCAACTTG GTTATCAAGC AGAAGGGCT GGCTGGAGAA ACATTTACTT AACTGGCGCA 7680  
CAAGAGCTAC GAGTAGGTAT TCAAGCTGGC GCGCCTAAAA CCGCATCGGC AGATGTCATC 7740  
AGTGAAATGG ACATGCCGAC TCTATTTGAC TTCCTCGCGG TGAAGATTGA TAGTCAACAG 7800  
GCGGCTAAGC ACGGCTTAGT TAAGATGAAT GTTATCACCC CTGATACTAA AGATATTCTC 7860  
TATATTGAGC TAAGCAACGG TAACTTAAGC AACGCAGTGG TCGACAAAAGA GCAAGCAGCT 7920  
GACGCAAACC TTATGGTTAA TAAAGCTGAC GTTAACCGCA TCTTACTTGG CCAAGTAACC 7980  
CTAAAAGCGT TATTAGCCAG CGGCGATGCC AAGCTCACTG GTGATAAAAC GGCATTTAGT 8040  
AAAATAGCCG ATAGCATGGT CGAGTTTACA CCTGACTTCG AAATCGTACC AACGCCCTGT 8100  
AAATGAGGCA TTAATCTCAA CAAGTGCAAG CTAGACATAA AAATGGGGCG ATTAGACGCC 8160

FIG. 4A-8

14/134

CCATTTTAA TGCAATTTTG AACTAGCTAG TCTTAGCTGA AGCTCGAACA ACAGCTTTAA 8220  
AATTCACCTC TTCTGCTGCA ATACTTATTT GCTGACACTG ACCAATACTC AGTGCAAAAC 8280  
GATAACTATC ATCAAGATGG CCCAGTAAAC AATGCCAATT ATCAGCAGCG TTCAATTTGCT 8340  
GTTCTTTAGC CTCGAATCAAA CCTAAACCAG ACTTTTGTGG CTCAGCCGTTA GGCTTATTAG 8400  
AACTCGACTC TAGTAAAGCA AGACCAATAT CTGTGTTTAA CAAAACCTGT CGCTGATTAA 8460  
GTTGATGCTC AACCTTGTGA TCCGCAATAG CATCGGAAAT ATCAACACAA TGGCTCAAGC 8520  
TTTTAGGTGC ATTAACTCCA AGAAAAGTTT CGCTCAGTGC AGAGAAAGTCA AACGCAAAAG 8580  
ATTTTAGCGA TAAATGCCAGC CCAAGTCCTT TCGCTTTAAT GTAAGACTCC TTGAGCGCCC 8640  
ACAAATCAAA AAAAGCGGTCT CGCTGCAAGG CCTCTGGTAA CGCTAACCAAG GCTCGCTTTT 8700  
CTGATTCAGA GAAATAATGA CTAAGAATAG AGTGGATATT GGTGCTGTTA CGGCAACGCT 8760  
CAATGTCGAC GCCAAACTCA ATACTAGCAG AGTCAGTTTC CTCCTTGCTT GCCTGACTGG 8820  
CGCCTTTATT ATCAGCAGTG CAAATGCCCTA CTAATAGCCA ATCTCCACTA TGAATCACAT 8880  
TAAAGTGGAC CCCGGTTTGA GCAAATTGCG CATCACTCAA TCTAGGCTTA CCTTTGTGCG 8940  
CATATTCAAA GGGCCATTCA TTGGGGCGTA TTTCACATAG TTGTGACAAAT AAAGCGCGCA 9000  
AATAGCCTCT TACCATTAAA CCTTGAGTTT TAGCTTCTTG TTAAATGTAG CGATTAACTT 9060  
TAATTAACTC ATCTTCAGGC AGCCATGACT TAACCAACTC TGTAAGTCTGG TTATCGCACT 9120  
CTTGATTTGT TAACGGACAG AAGTATAAGG AAATCAATCG AGAAGTTAGC AATTTTTCAG 9180

FIG. 4A-9

15/134

GACACTCTTT AAAGCAACAA ACATAACCCC TATTTTACC AATTTAAGAT CAAAACTAAA 9240  
GCCAAACTA ATTGAGAATA GTGTCAAACT AGCTTTAAAG GAAAAAATA TAAAAAGAAC 9300  
ATTATACTTG TATAAATTAT TTTACACACC AAAGCCATGA TCTTCACAA ATTAGCTCCC 9360  
TCTCCCTAAA ACAAGATTGA ATAAAAAAAT AAACCTTAAC TTTCATATAG ATAAAACAAA 9420  
CCAATGGGAT AAAGTATATT GAATTCATTT TTAAGGAAAA ATTCAAATTG AATTCAGCT 9480  
CTTCAGTAAA AGCATATTTT GCCGTTAGTG TGAATAAAAA CAAATTTAAA AACCAACATA 9540  
GAACAAATA GCAGACAATA AAACCAAGGC GCAACACAAA CAACGCGCTT ACAATTTTCA 9600  
CAAAAAAGCA ACAAGAGTAA CGTTTAGTAT TTGGATATGG TTATTGTAAT TGAGAATTTT 9660  
ATAACAATTA TATTAAGGA ATGAGTATGT TTTTAAATC AAAACTTTTCG CGCTCAGTCA 9720  
AACTTGCCAT ATCCGCAGGC TTAACAGCCT CGCTAGCTAT GCCTGTTTTT GCAGAAGAAA 9780  
CTGCTGCTGA AGAACAAATA GAAAGAGTCG CAGTGACCGG ATCGCGAATC GCTAAAGCAG 9840  
AGCTAACTCA ACCAGCTCCA GTCGTCAGCC TTTCAGCCGA AGAACTGACA AAATTTGGTA 9900  
ATCAAGATTT AGGTAGCGTA CTAGCAGAAT TACCTGCTAT TGGTGCAACC AACACTATTA 9960  
TTGGTAATAA CAATAGCAAC TCAAGCGCAG GTGTTAGCTC AGCAGACTTG CGTCGTCTAG 10020  
GTGCTAACAG AACCTTAGTA TTAGTCAACG GTAAGCGCTA CGTTGCCGC CAACCGGGCT 10080  
CAGCTGAGGT AGATTTGTCA ACTATACCAA CTAGCATGAT CTCGCGAGTT GAGATTGTAA 10140  
CCGGCGGTGC TTCAGCAATT TATGTTCCG ACGCTGTATC AGGTGTTATC AACGTTATCC 10200

FIG. 4A-10



16/134

TTAAAGAAGA CTTTGAAGGC TTTGAGTTA ACGCAGTAC TAGCGTTCT ACTGAAAGTG 10260  
TAGGCACTCA AGAGCACTCT TTTGACATTT TGGGTGGTGC AAACGTTGCA GATGGACGTG 10320  
GTAATGTAAC CTTCTACGCA GGTATGAAC GTACAAAAGA AGTCATGGCT ACCGACATTC 10380  
GCCAATTGCA TGCTTGGGGA ACAATTAAA ACGAAGCCGA TGGTGGTGAA GATGATGGA 10440  
TTCCAGACAG ACTACGTGTA CCACGAGTTT ATTCTGAAAT GATTAATGCT ACCGGTGTTA 10500  
TCAATGCATT TGGTGGTGA ATTGGTCGCT CAACCTTTGA CAGTAAACGC ATCCTATTG 10560  
CACAAACAAGA ACGTGATGGG ACTAACAGCT TTGCATTTGG TTCATTCCCT AATGGCTGTG 10620  
ACACATGTTT CAACACTGAA GCATACGAAA ACTATATTCC AGGGGTAGAA AGAATAAACG 10680  
TTGGCTCATC ATTCAACTTT GATTTTACCG ATAACATTCA ATTTTACACT GACTTCAGAT 10740  
ATGTAAAGTC AGATAATTCAG CAACAATTC AGCCTTCATT CCGTTTGGT AACATTAATA 10800  
TCAAATGTTGA AGATAACGCC TTTTGTGAATG ACGACTTGGC TCAGCAAAATG CTCGATGCGG 10860  
GTCAAACCAA TGCTAGTTT GCCAAGTTT TTGATGAATT AGGAAATCGC TCAGCAGAAA 10920  
ATAAACGCGA ACTTTTCCGT TACGTAGGTG GCTTAAAGG TGGCTTTGAT ATTAGCGAAA 10980  
CCATATTGA TTACGACCCT TACTATGTTT ATGGCGAGAC TAATAACCGT CGTAAACCC 11040  
TTAATGACCT AATTCCCTGAT AACTTTGTG CAGCTGTGCA CTCTGTTATT GATCCTGATA 11100  
CTGGCTTAGC AGCGTGTGCG TCACAAGTAG CAAGCGCTCA AGCGGATGAC TATACAGATC 11160  
CCGCGTCTGT AAATGGTAGC GACTGTGTTG CTTATAACCC ATTGGCATG GGTCAGGCTT 11220

FIG. 4A-11

17/134

CAGCAGAAGC CCGCGACTGG GTTCTGCTG ATGTGACTCG TGAAGACAAA ATAACTCAAC 11280  
AAGTGATTGG TGGTACTCTC GGTACCGATT CTGAAGAACT ATTTGAGCTT CAAGGTGGTG 11340  
CAATCGCTAT GGTGTGTTGGT TTTGAATACC GTGAAGAAAC GTCTGGTTCA ACAACCGATG 11400  
AATTTACTAA AGCAGGTTTC TTGACAAGCG CTGCAACGCC AGATTCTTAT GGCGAATACG 11460  
ACGTGACTGA GTATTTTGGT GAGGTGAACA TCCCAGTACT AAAAGAATTA CCTTTGCAC 11520  
ATGAGTTGAG CTTTGACGGT GCATACCGTA ATGCTGATTA CTCACATGCC GGTAAGACTG 11580  
AAGCATGGAA AGCTGGTATG TTCTACTCAC CATTAGAGCA ACTTGCATTA CGTGGTACGG 11640  
TAGGTGAAGC AGTACGAGCA CCAACATTG CAGAAAGCCTT TAGTCCACGC TCTCCTGGTT 11700  
TTGGCCGCCGT TTCAGATCCA TGTGATGCAG ATAACATTAA TGACGATCCG GATCGCGTGT 11760  
CAAACGTGTC AGCATTGGGG ATCCCTCCAG GATTCCAAGC TAATGATAAC GTCAGTGTAG 11820  
ATACCTTATC TGGTGGTAAC CCAGATCTAA AACCTGAAAC ATCAACATCC TTTACAGGTG 11880  
GTCTTGTTG GACACCAACG TTTGCTGACA ATCTATCATT CACTGTCGAT TATTATGATA 11940  
TTCAAATTGA GGATGCTATT TTGTCAGTAG CCACCCAGAC TGTGGCTGAT AACTGTGTTG 12000  
ACTCAACTGG CGGACCTGAC ACCGACTTCT GTAGTCAAGT TGATCGTAAT CCAACGACCT 12060  
ATGATATTGA ACTTGTTGCG TCTGGTTATC TAAATGCCG GCATTGAAT ACCAAAGGTA 12120  
TTGAATTTCA AGCTGCATAC TCATTAGATC TAGAGTCTTT CAACGCGCCT GGTGAACACT 12180  
GCTTCAACCT ATTGGGGAAC CAATTACTTG AACTAGAACG TCTTGAATTC CAAAATCGTC 12240

FIG. 4A-12

18/134

CTGATGAGAT TAAATGATGAA AAAGGCGAAG TAGGTGATCC AGAGCTGCAG TTCCGCCTAG 12300  
GCATCGATTA CCGTCTAGAT GATCTAAGTG TTAGCTGGAA CACGCGTTAT ATTGATAGCG 12360  
TAGTAACTTA TGATGTCTCT GAAAATGGTG GCTCTCCTGA AGATTATAT CCAGGCCACA 12420  
TAGGCTCAAT GACAACTCAT GACTTGAGCG CTACATACTA CATCAATGAG AACTTCATGA 12480  
TTAACGGTGG TGTACGTAAC CTATTTGACG CACTTCCACC TGGATACACT AACGATGCGC 12540  
TATATGATCT AGTTGGTCGC CGTGCATTCC TAGGTATTAA GGTAATGATG TAATTAATTA 12600  
TTACGCCCTCT AACTAATAAA AATGCAATCT CTTCGTAGAG ATTGCATTTT TTTATGAAAT 12660  
CCAATCTTAA ACTGGTTCTC CGAGCATCTT ACGCCTTAAA AACCCCGCCC CTCAATGTAA 12720  
CGCCAAAGTT AATTGCTTAC ACGCACTTAC ACAACGAAAC AATTTCATTA ACACGAGACA 12780  
CAGCTCACGC TTTTATTATT ACCCTTGATT TTA CTACATA AAATTGCGTT TTAGCGCACA 12840  
AGTGTCTCC CAAGCTGGTC GTATCTGTAA TTATTCAGTC CCAGGTGATT GTATTGACCC 12900  
ATAAGCTCAG GTAGTCTGCT CTGCCATTAG CTAAACAATA TTGACAAAAAT GCGGATAAAA 12960  
TGTGGCTTAG CGCTAAGTTC ACCGTAAGTT TTATCGGCAT TAAGTCCCAA CAGATTATTA 13020  
ACGGAAAACC GCTAAACTGA TGGCAAAAAT AAATAGTGAA CACTTGGATG AAGCTACTAT 13080  
TACTTCGAAT AAGTGTAACG AACACAGAGAC TGAGGCTCGG CATAGAAATG CCACTACAAC 13140  
ACCTGAGATG CGCCGATTCA TACAAGAGTC GGATCTCAGT GTTAGCCAAC TGTCTAAAAAT 13200  
ATTAAATATC AGTGAAGCTA CCGTACGTAA GTGGCGCAAG CGTGACTCTG TCGAAAACTG 13260

FIG. 4A-13

19/134

TCCTAATACC CCGCACCATC TCAATACCAC GCTAACCCCT TTGCAAGAAT ATGTGGTTGT 13320  
GGGCCTGCGT TATCAATTGA AAATGCCATT AGACAGATTG CTCAAAGCAA CCCAAGAGTT 13380  
TATCAATCCA AACGTGTCGC GCTCAGGTTT AGCAAGATGT TTGAAGCGTT ATGGCGTTTC 13440  
ACGGGTGAGT GATATCCAAA GCCCACACGT ACCAATGCGC TACTTTAATC AAATTCCAGT 13500  
CACTCAAGGC AGCGATGTGC AAACCTACAC CCTGCACAT GAAACGCTGG CAAAAACCTT 13560  
AGCCTTACCT AGTACCGATG GTGACAATGT GGTGCAAGTG GTGTCTCTCA CCATTCCACC 13620  
AAAGTTAACC GAAGAAGCAC CCAGTTCAAT TTTGCTCGGC ATTGATCCTC ATAGCGACTG 13680  
GATCTATCTC GACATATACC AAGATGGCAA TACACAAGCC ACGAATAGAT ATATGGCTTA 13740  
TGTGCTAAAA CACGGGCCAT TCCATTACG AAAGTTACTC GTGCGTAACT ATCACACCTT 13800  
TTTACAGCGC TTTCCCTGGAG CGACGCAAAA TCGCCGCCCC TCTAAAGATA TGCCTGAAAC 13860  
AATCAACAAG ACGCCTGAAA CACAGGCACC CAGTGGAGAC TCATAATGAG CCAGACCTCT 13920  
AAACCTACAA ACTCAGCAAC TGAGCAAGCA CAAGACTCAC AAGCTGACTC TCGTTTAAAT 13980  
AAACGACTAA AAGATATGCC AATTGCTATT GTTGGCATGG CGAGTATTTT TGCAAAACTCT 14040  
CGCTATTGTA ATAAGTTTGG GGACTTAATC AGCGAAAAAA TTGATGCGAT TACTGAATTA 14100  
CCATCAACTC ACTGGCAGCC TGAAGAATAT TACGAGCGAG ATAAAAACCGC AGCAGACAAA 14160  
AGCTACTGTA AACGTGGTGG CTTTTTGCCA GATGTAGACT TCAACCCCAAT GGAGTTTGGC 14220  
CTGCCGCCAA ACATTTTGGG ACTGACCGAT TCATCGCAAC TATTATCACT CATCGTTGCT 14280

FIG. 4A-14

20/134

AAAGAAGTGT TGGCTGATGC TAACTTACCT GAGAATTACG ACCGCGATAA AATTGGTATC 14340  
ACCTTAGGTG TCGGCGGTGG TCAAAAAATT AGCCACAGCC TAACAGCGCG TCTGCAATAC 14400  
CCAGTATTGA AGAAAGTATT CGCCAATAGC GGCATTAGTG ACACCGACAG CGAATATGCTT 14460  
ATCAAGAAAT TCCAAGACCA ATATGTACAC TGGGAAGAAA ACTCGTTCCC AGGTTCACTT 14520  
GGTAACGTTA TTGCGGGCCG TATCGCCAAC CGCTTCGATT TTGGCGGCAT GAACTGTGTG 14580  
GTTGATGCTG CCTGTGCTGG ATCACTTGCT GCTATGCGTA TGGCGCTAAC AGAGCTAACT 14640  
GAAGGTCGCT CTGAAATGAT GATCACCGGT GGTGTGTGTA CTGATAACTC ACCCTCTATG 14700  
TATATGAGCT TTTCAAAAAC GCCCGCCTTT ACCACTAACG AAACCATTC A GCCATTTGAT 14760  
ATCGACTCAA AAGGCATGAT GATTGGTGAA GGTATTGGCA TGGTGGCGCT AAAGCGTCTT 14820  
GAAGATGCAG AGCGCGATGG CGACCGCATT TACTCTGTAA TTAAGAAGTGT GGGTGCATCA 14880  
TCTGACGGTA AGTTTAAATC AATCTATGCC CCTCGGCCAT CAGGCCAAGC TAAAGCACTT 14940  
AACCGTGCCT ATGATGACGC AGGTTTTCGG CCGCATACCT TAGGTCTAAT TGAAGCTCAC 15000  
GGAACAGGTA CTGCAGCAGG TGACGCGGCA GAGTTTGCCG GCCTTTGCTC AGTATTGTCT 15060  
GAAGGCAACG ATACCAAGCA ACACATTGCG CTAGGTTTCAG TTAATATCACA AATTGGTCAT 15120  
ACTAAATCAA CTGCAGGTAC AGCAGGTTTA ATTAAAGCTG CTCTTGCTTT GCATCACAAAG 15180  
GTACTGCCGC CGACCATTA CGTTAGTCAG CCAAGCCCTA AACTTGATAT CGAAAACTCA 15240  
CCGTTTATC TAAACACTGA GACTCGTCCA TGGTTACCAC GTGTTGATGG TACGCCGCGC 15300

FIG. 4A-15

21/134

CGCGCGGGTA TTAGCTCATT TGGTTTGGT GGCACTAACT TCATTTTGT ACTAGAAAGAG 15360  
TACAACCAAG AACACAGCCG TACTGATAGC GAAAAAGCTA AGTATCGTCA ACGCCAAAGTG 15420  
GCGCAAAGCT TCCTTGTTAG CGCAAGCGAT AAAGCATCGC TAATTAACGA GTTAAACGTA 15480  
CTAGCAGCAT CTGCAAGCCA AGCTGAGTTT ATCCTCAAAG ATGCAGCAGC AAACATATGGC 15540  
GTACGTGAGC TTGATAAAAA TGCACCACGG ATCGGTTTAG TTGCAAAACAC AGCTGAAGAG 15600  
TTAGCAGGCC TAATTAAGCA AGCACTTGCC AAAC TAGCAG CTAGCGATGA TAACGCATGG 15660  
CAGCTACCTG GTGGCACTAG CTACCGCGCC GCTGCAGTAG AAGTAAAGT TGCCGCACTG 15720  
TTTGCTGGCC AAGGTTTACA ATATCTCAAT ATGGGCCGTG ACCTTACTTG TTATTACCCA 15780  
GAGATGCGTC AGCAATTGT AACTGCAGAT AAAGTATTG CCGCAAAATGA TAAAACGCCG 15840  
TTATCGCAAA CTCTGTATCC AAAGCCTGTA TTTAATAAAG ATGAATTAAA GGCTCAAGAA 15900  
GCCATTTTGA CCAATACCGC CAATGCCCAA AGCGGAATTG GTGCGATTTC AATGGGTCAA 15960  
TACGATTTGT TTAGTGCGGC TGGCTTTAAT GCCGACATGG TTGCAGGCCA TAGCTTTGGT 16020  
GAGCTAAGTG CACTGTGTGC TGCAGGTGTT ATTTCAGCTG ATGACTACTA CAAGCTGGCT 16080  
TTTGCTCGTG GTGAGGCTAT GGCAACAAAA GCACCGGCTA AAGACGGCGT TGAAGCAGAT 16140  
GCAGGAGCAA TGTTTGCAAT CATAACCAAG AGTGCTGCAG ACCTTGAAAC CGTTGAAGCC 16200  
ACCATCGCTA AATTGTATGG GGTGAAAGTC GCTAACTATA ACGCGCCAAAC GCAATCAGTA 16260  
ATTGCAGGCC CAACAGCAAC TACCGCTGAT GCGGCTAAAG CGCTAACTGA GCTTGGTTAC 16320

FIG. 4A-16

22/134

AAAGCGATTA ACCTGCCAGT ATCAGGTGCA TTCCAQACTG AACTTGTTGG TCACGCTCAA 16380  
GGCCCATTTG CTAAAGCGAT TGACGCAGCC AAATTACTA AAACAAGCCG AGCACTTTAC' 16440  
TCAAATGCAA CTGGCGGACT TTATGAAAGC ACTGCTGCAA AGATTAAAGC CTCGTTTAAG 16500  
AAACATATGC TTCAATCAGT GCGCTTTACT AGCCAGCTAG AAGCCATGTA CAACGACGGC 16560  
GCCCGTGTAT TTGTTGAATT TGGTCCAAAG AACATCTTAC AAAAATTAGT TCAAGGCACG 16620  
CTTGTCACAA CTGAAAAATGA AGTTTGCACT ATCTCTATCA ACCCTAATCC TAAAGTTGAT 16680  
AGTGATCTGC AGCTTAAAGCA AGCAGCAATG CAGCTAGCGG TTACTGGTGT GGTACTCAGT 16740  
GAAATTGACC CATACCAAGC CGATATTGCC GCACCAGCGA AAAAGTCGCC AATGAGCATT 16800  
TCGCTTAATG CTGCTAACCA TATCAGCAAA GCAACTCGCG CTAAGATGGC CAAGTCCTTA 16860  
GAGACAGGTA TCGTCACCTC GCAAAATAGAA CATGTTATTG AAGAAAAAAT CGTTGAAGTT 16920  
GAGAAACTGG TTGAAAGTCGA AAAGATCGTC GAAAAAGTGG TTGAAGTAGA GAAAGTTGTT 16980  
GAGGTTGAAG CTCCTGTTAA TTCAGTGCAA GCCAATGCAA TTCAAACCCG TTCAGTTGTC 17040  
GCTCCAGTAA TAGAGAACCA AGTCGTGTCT AAAAACAGTA AGCCAGCAGT CCAGAGCATT 17100  
AGTGGTGATG CACTCAGCAA CTTTTTTGCT GCACAGCAGC AAACCGCACA GTTGCAATCAG 17160  
CAGTTCTTAG CTATTCCGCA GCAATATGGT GAGACGTTCA CTACGCTGAT GACCGAGCAA 17220  
GCTAAACTGG CAAGTTCTGG TGTGCAATT CCAGAGAGTC TGCAACGCTC AATGGAGCAA 17280  
TTCCACCAAC TACAAGCGCA AACACTACAA AGCCACACCC AGTTCCTTGA GATGCAAGCG 17340

FIG. 4A-17

23/134

GGTAGCAACA TTGCAGCGTT AACCTACTC AATAGCAGCC AAGCAACTTA CGCTCCAGCC 17400  
ATTCACAATG AAGCGATTCA AAGCCAAAGTG GTTCAAAGCC AAAGTGCAGT CCAGCCAGTA 17460  
ATTTCAACAC AAGTTAACCA TGTGTCAGAG CAGCCAACTC AAGCTCCAGC TCCAAAAGCG 17520  
CAGCCAGCAC CTGTGACAAC TGCAGTTCAA ACTGCTCCGG CACAAGTTGT TCGTCAAGCC 17580  
GCACCAGTTC AAGCCGCTAT TGAACCGATT AATACAAAGTG TTGCGACTAC AACGCCTTCA 17640  
GCCTTCAGCG CCGAAACAGC CCTGAGCGCA ACAAAGTCC AAGCCACTAT GCTTGAAAGTG 17700  
GTTGCTGAGA AAACCGGTTA CCCAACTGAA ATGCTAGAGC TTGAAAATGGA TATGGAAAGCC 17760  
GATTTAGGCA TCGATTCTAT CAAGCGTGTA GAAATTCTTG GCACAGTACA AGATGAGCTA 17820  
CCGGGTCTAC CTGAGCTTAG CCTGAAGAT CTAGCTGAGT GTCGAACGCT AGGCGAAATC 17880  
GTTGACTATA TGGGCAGTAA ACTGCCGGCT GAAGGCTCTA TGAATTCTCA GCTGTCTACA 17940  
GGTTCCGCAG CTGCGACTCC TGCAGCGAAT GGTCTTTCTG CGGAGAAAGT TCAAGCGACT 18000  
ATGATGTCTG TGGTTGCCGA AAAGACTGGC TACCCCAACTG AAATGCTAGA GCTTGAAATG 18060  
GATATGGAAG CCGATTTAGG CATAGATTCT ATCAAGCGCG TTGAAATTCT TGGCACAGTA 18120  
CAAGATGAGC TACCGGGTCT ACCTGAGCTT AGCCCTGAAG ATCTAGCTGA GTGTCGTACT 18180  
CTAGGCGAAA TCGTTGACTA TATGAACTCT AAAGTCCGCTG ACGGCTCTAA GCTGCCGGCT 18240  
GAAGGCTCTA TGAATTCTCA GCTGTCTACA AGTGCCGCAG CTGCGACTCC TGCAGCGAAT 18300  
GGTCTCTCTG CGGAGAAAGT TCAAGCGACT ATGATGTCTG TGGTTGCCGA AAAGACTGGC 18360

FIG. 4A-18



24/134

TACCCAACTG AAATGCTAGA ACTTGAAATG GATATGGAAG CTGACCTTGG CATCGATTCA 18420  
ATCAAGCGCG TTGAAATTTCT TGGCACAGTA CAAGATGAGC TACCGGGTTT ACCTGAGCTA' 18480  
AATCCAGAAG ATTTGGCAGA GTGTCGTACT CTTGGCGAAA TCGTGACTTA TATGAACTCT 18540  
AAACTCGCTG ACGGCTCTAA GCTGCCAGCT GAAGGCTCTA TGCACTATCA GCTGTCTACA 18600  
AGTACCGCTG CTGCGACTCC TGTAGCGAAT GGTCCTCTCTG CAGAAAAAGT TCAAGCGACC 18660  
ATGATGCTG TAGTTGCAGA TAAAACTGGC TACCCAACTG AAATGCTTGA ACTTGAAATG 18720  
GATATGGAAG CCGATTTAGG TATCGATTCT ATCAAGCGCG TTGAAATTCT TGGCACAGTA 18780  
CAAGATGAGC TACCGGGTTT ACCTGAGCTA AATCCAGAAG ATCTAGCAGA GTGTCGCACC 18840  
CTAGGCGAAA TCGTTGACTA TATGGGCAGT AAATGCGCG CTGAAGGCTC TGCTAATACA 18900  
AGTGCCGCTG CGTCTCTTAA TGTTAGTCC GTTGCGGCGC CTCGAAGCTGC TGCGACTCCT 18960  
GTATCGAACG GTCTCTCTGC AGAGAAAGTG CAAAGGACTA TGATGTCAGT AGTTGCAGAA 19020  
AAGACCGGCT ACCCAACTGA AATGCTAGAA CTTGGCATGG ATATGGAAGC CGATTTAGGT 19080  
ATCGACTCAA TTAAACGCGT TGAGATTCTT GGCACAGTAC AAGATGAGCT ACCGGGTCTA 19140  
CCAGAGCTTA ATCCTGAAGA TTTAGCTGAG TGCCGTACGC TGGGCGAAAT CGTTGACTAT 19200  
ATGAACTCTA AGCTGGCTGA CGGCTCTAAG CTTCCAGCTG AAGGCTCTGC TAATACAAGT 19260  
GCCACTGCTG CGACTCCTGC AGTGAATGGT CTTTCTGCTG ACAAGGTACA GCGGACTATG 19320  
ATGTCTGTAG TTGCTGAAAA GACCGGCTAC CCAACTGAAA TGCTAGAACT TGGCATGGAT 19380

FIG. 4A-19

25/134

ATGGAAGCAG ACCTTGGTAT TGATTCTATT AAGCGCGTTG AAATTCTTGG CACAGTACAA 19440  
GATGAGCTCC CAGGTTTACC TGAGCTTAAT CCTGAAGATC TCGCTGAGTG CCGCACGCTT 19500  
GGCGAAATCG TTAGCTATAT GAACTCTCAA CTGGCTGATG GCTCTAAACT TTCTACAAAGT 19560  
GCGGCTGAAG GCTCTGCTGA TACAAAGTGCT GCAAATGCTG CAAAGCCGGC AGCAATTTTCG 19620  
GCAGAACCAA GTGTTGAGCT TCCTCCTCAT AGCGAGGTAG CGCTAAAAAA GCTTAATGCG 19680  
GCGAACCAAGC TAGAAAAATTG TTTCGCCGCA GACGCAAGTG TTGTGATTAA CGATGATGGT 19740  
CACAAACGCAG GCGTTTTAGC TGAGAAAACTT ATTAAACAAG GCCTAAAAAGT AGCCGTTGTG 19800  
CGTTTACCGA AAGGTCAGCC TCAATCGCCA CTTTCAAGCG ATGTTGCTAG CTTTGAGCTT 19860  
GCCTCAAGCC AAGAACTCTGA GCTTGAAAGCC AGTATCACTG CAGTTATCGC GCAGATTGAA 19920  
ACTCAGGTTG GCGCTATTGG TGGCTTTATT CACTTGCAAC CAGAAAGCGAA TACAGAAAGAG 19980  
CAAACGGCAG TAAACCTAGA TCGGCAAAAGT TTTACTCAGC TTAGCAATGC GTTCTTGTGG 20040  
GCCAAATTAT TGCAACCCAAA GCTCGTTGCT GGAGCAGATG CGCGTCGCTG TTTTGTAAACA 20100  
GTAAGCCGTA TCGACGGTGG CTTTGGTTAC CTAAATACTG ACGCCCTAAA AGATGCTGAG 20160  
CTAAACCAAG CAGCATTAGC TGGTTTAACT AAAACCTTAA GCCATGAATG GCCACAAAGTG 20220  
TTCTGTGCGG CGCTAGATAT TGCAACAGAT GTTGATGCAA CCCATCTTGC TGATGCAATC 20280  
ACCAGTGAAC TATTTGATAG CCAAGCTCAG CTACCTGAAG TGGGCTTAAG CTTAAATTGAT 20340  
GGCAAAAGTTA ACCGCGTAAC TCTAGTTGCT GCTGAAGCTG CAGATAAAAC AGCAAAAGCA 20400

FIG. 4A-20

26/134

GAGCTTAACA GCACAGATAA AATCTTAGTG ACTGGTGGGG CAAAAGGGGT GACATTTGAA 20460  
TGTGCACTGG CATTAGCATC TCGCAGCCAG TCTCACTTTA TCTTAGCTGG GCGCAGTGAA 20520  
TTACAAGCTT TACCAAGCTG GGCTGAGGGT AAGCAAACTA GCGAGCTAAA ATCAGCTGCA 20580  
ATCGCACATA TTATTTCTAC TGGTCAAAAG CCAACGCCTA AGCAAGTTGA AGCCGCTGTG 20640  
TGGCCAGTGC AAAGCAGCAT TGAATAAT GCGGCCCTAG CCGCCTTTAA CAAAGTTGGC 20700  
GCCTCAGCTG AATACGTCAG CATGGATGTT ACCGATAGCG CCGCAATCAC AGCAGCACTT 20760  
AATGGTCGCT CAAATGAGAT CACCGGTCTT ATTCAATGGC CAGGTGTACT AGCCGACAAG 20820  
CATATTCAAG ACAAGACTCT TGCTGAACTT GCTAAAGTTT ATGGCACTAA AGTCAACGGC 20880  
CTAAAAGCGC TGCTCGCGGC ACTTGAGCCA AGCAAAATTA AATTACTTGC TATGTTTCTCA 20940  
TCTGCAGCAG GTTTTACGG TAATATCGGC CAAAGCGATT ACGCGATGTC GAACGATATT 21000  
CTTAACAAGG CAGCGCTGCA GTTCACCGCT CGCAACCCAC AAGCTAAAAGT CATGAGCTTT 21060  
AACTGGGGTC CTTGGGATGG CGGCATGGTT AACCCAGCGC TTAAAAAGAT GTTACCCGAG 21120  
CGTGGTGTGT ACGTTATTCC ACTAAAAGCA GGTGCAGAGC TATTTGCCAC TCAGCTATTG 21180  
GCTGAAACTG GCGTGCAGTT GCTCATTTGGT ACGTCAATGC AAGGTGGCAG CGACACTAAA 21240  
GCAACTGAGA CTGCTTCTGT AAAAAAGCTT AATCGGGTG AGGTGCTAAG TGCATCGCAT 21300  
CCGCGTGTG GTGCACAAAA AACACCACTA CAAGCTGTCA CTGCAACGCG TCTGTTAACC 21360  
CCAAGTGCCA TGGTCTTTCAT TGAAGATCAC CGCATTTGGC GTAACAGTGT GTTGCCAACG 21420

FIG. 4A-21

27/134

GTATGCGCCA TCGACTGGAT GCGTGAAGCG GCAAGCGACA TGCTTGGCGC TCAAGTTAAG 21480  
GTAAGTTGATT ACAAGCTATT AAAAGGCATT GTATTTGAGA CTGATGAGCC GCAAGAGTTA 21540  
ACACTTGAGC TAACGCCAGA CGATTACAGC GAAGCTACGC TACAAGCATT AATCAGCTGT 21600  
AATGGGCGTC CGCAATACAA GCGACGCTT ATCAGTGATA ATGCCGATAT TAAGCAACTT 21660  
AACAAGCAGT TTGATTTAAG CGCTAAGCG ATTACCACAG CAAAAGAGCT TTATAGCAAC 21720  
GGCACCTTGT TCCACGGTCC GCGTCTACAA GGGATCCCAAT CTGTAGTGCA GTTCGATGAT 21780  
CAAGGCTTAA TTGCTAAAAGT CGCTCTGCCT AAGGTTGAAC TTAGCGATTG TGGTGAGTTC 21840  
TTGCCGCAAA CCCACATGGG TGGCAGTCAA CCTTTTGCTG AGGACTTGCT ATTACAAGCT 21900  
ATGCTGGTTT GGGCTCGCCT TAAAACTGGC TCGGCAAGTT TGCCATCAAG CATTGGTGAG 21960  
TTTACCTCAT ACCAACCCTT GGCCTTTGGT GAAACTGGTA CCATAGAGCT TGAAGTGATT 22020  
AAGCACAAAC AACGCTCACT TGAAGCGAAT GTTGCCTAT ATCGTGACAA CGGCGAGTTA 22080  
AGTGCCATGT TTAAGTCAGC TAAATCACC ATTAGCAAAA GCTTAAATTC AGCATTTTAA 22140  
CCTGCTGTCT TAGCAAAACGA CAGTGAGGCG AATTAGTGA ACAACGCCT AAAGCTAGTG 22200  
CGATGCCGCT GCGCATCGCA CTTATCTTAC TGCCAACACC GCAGTTTGAA GTTAACTCTG 22260  
TCGACCCAGTC AGTATTAGCC AGCTATCATA CACTGCAGCC TGAGCTAAAT GCCCTGCTTA 22320  
ATAGTGCGCC GACACCTGAA ATGCTCAGCA TCACTATCTC AGATGATAGC GATGCAACA 22380  
GCTTTGAGTC GCAGCTAAAT GCTGCGACCA ACGCAATTAA CAATGGCTAT ATCGTCAAAGC 22440

FIG. 4A-22

28/134

TTGCTACGGC AACTCACGCT TTGTTAATGC TGCTGCATT AAAAGCGCG CAAATGCGGA 22500  
TCCATCCTCA TCGCGAGCTT GCCGCTATGC AGCAAGCTAA ATCGACGCCA ATGAGTCAAG '22560  
TATCTGGTGA GCTAAAGCTT GCGGCTAATG CGCTAAGCCT AGCTCAGACT AATGCGCTGT 22620  
CTCATGCTTT AAGCCAAGCC AAGCGTAACT TAACTGATGT CAGCGTGAAT GAGTGTTTTG 22680  
AGAACCTCAA AAGTGAACAG CAGTTCACAG AGGTTTATTC GCTTATTCAG CAACTTGCTA 22740  
GCCGCACCCA TGTGAGAAA GAGGTTAATC AAGGTGTGGA ACTTGGCCCT AAACAAGCCA 22800  
AAAGCCACTA TTGGTTTAGC GAATTCACC AAAACCGTGT TGCTGCCATC AACTTTATTA 22860  
ATGGCCCAACA AGCAACCAGC TATGTGCTTA CTCAAGGTTT AGGATTGTTA GCTGCGAAAT 22920  
CAATGCTAAA CCAGCAAAGA TTAATGTTTA TCTTGCCGGG TAACAGTCAG CAACAAATAA 22980  
CCGCATCAAT AACTCAGTTA ATGCAGCAAT TAGAGCGTTT GCAGGTAACT GAGGTTAATG 23040  
AGCTTTCTCT AGAATGCCAA CTAGAGCTGC TCAGGATAAT GTATGACAAC TTAGTCAACG 23100  
CAGACAAACT CACTACTCGC GATAGTAAGC CCGCTTATCA GGCTGTGATT CAAGCAAGCT 23160  
CTGTTAGCGC TGCAAAGCAA GAGTTAAGCG CGCTTAACGA TGCACTCACA GCGCTGTTTG 23220  
CTGAGCAAAC AAACGCCACA TCAACGAATA AAGGCTTAAT CCAATACAAA ACACCGCGG 23280  
GCAGTTACTT AACCTTAACA CCGCTTGGA GCAACAATGA CAACGCCCAA GCGGCTCTTG 23340  
CTTTTGCTA TCCGGGTGTG GGAACGGTTT ACGCCGATAT GCTTAATGAG CTGCATCAGT 23400  
ACTTCCCTGC GCTTTACGCC AAACCTGAGC GTGAAGCGGA TTTAAAGGCG ATGCTACAAG 23460

FIG. 4A-23

29/134

CAGAAGATAT CTATCATCTT GACCCTAAAC ATGCTGCCCA AATGAGCTTA GGTGACTTAG 23520  
CCATTGCTGG CGTGGGAGC AGCTACCTGT TAACTCAGCT GCTCACCAGT GAGTTTAATA 23580  
TTAAGCCTAA TTTTGCATTA GGTACTCAA TGGGTGAAGC ATCAATGTGG GCAAGCTTAG 23640  
GCGTATGGCA AAACCCGCAT GCGCTGATCA GCAAAACCCA AACCGACCCG CTATTACTT 23700  
CTGCTATTTC CGGCAAATTG ACCGCGGTTA GACAAGCTTG GCAGCTTGAT GATACCGCAG 23760  
CGGAAATCCA GTGGAATAGC TTTGTGGTTA GAAGTGAAGC AGCGCCGATT GAAAGCCTTG 23820  
TAAAAGATTA CCCACACGCT TACCTGCGA TTATTCAAGG GGATACCTGC GTAAATCGCTG 23880  
GCTGTGAAAT CCAATGTAAA GCGCTACTTG CAGCACTGGG TAAACGCGGT ATTGCAGCTA 23940  
ATCGTGTAAAC GCGGATGCAT ACGCAGCCTG CGATGCAAGA GCATCAAAAT GTGATGGATT 24000  
TTTATCTGCA ACCGTTAAAA GCAGAGCTTC CTAGTGAAAT AAGCTTTATC AGCGCCGCTG 24060  
ATTTAACTGC CAAGCAAACG GTGAGTGAGC AAGCACTTAG CAGCCAAGTC GTTGCTCAGT 24120  
CTATTGCCGA CACCTTCTGC CAAACCTTGG ACTTTACCGC GCTAGTACAT CACGCCCAAC 24180  
ATCAAGGCGC TAAGCTGTTT GTTGAAATTG GCGCGGATAG ACAAAACTGC ACCTTGATAG 24240  
ACAAGATTGT TAAACAAGAT GGTGCCAGCA GTGTACAACA TCAACCTTGT TGCACAGTGC 24300  
CTATGAACGC AAAAGGTAGC CAAGATATTA CCAGCGTGAT TAAAGCGCTT GGCCAATTAA 24360  
TTAGCCATCA GGTGCCATTA TCGGTGCAAC CATTATTGA TGGACTCAAG CGCGAGCTAA 24420  
CACTTTGCCA ATTGACCAGC CAACAGCTGG CAGCACATGC AAATGTTGAC AGCAAGTTTG 24480

FIG. 4A-24

30/134

AGTCTAACCA AGACCATTTA CTTCAAGGGG AAGTCTAATG TCATTACCAG ACAATGCTTC 24540  
TAACCACTT TCTGCCAACC AGAAAGGCGC ATCTCAGGCA AGTAAAAACCA GTAAGCAAAG' 24600  
CAAAATCGCC ATTGTCGGTT TAGCCACTCT GTATCCAGAC GCTAAAAACCC CGCAAGAATT 24660  
TTGGCAGAAT TTGCTGGATA AACGGGACTC TCGCAGCACC TTAACCTAACG AAAAACTCGG 24720  
CGCTAACAGC CAAGATTATC AAGGTGTGCA AGGCCAATCT GACCGTTTTT ATTGTAATAA 24780  
AGGCGGCTAC ATTGAGAACT TCAGCTTTAA TGCTGCAGGC TACAAATGTC CGGAGCAAAG 24840  
CTTAAATGGC TTGGACGACA GCTTCCTTTG GGGCTCGAT ACTAGCCGTA ACGACTAAT 24900  
TGATGCTGGT ATTGATATCA ACGGCGCTGA TTTAAGCCGC GCAGGTGTAG TCATGGGCGC 24960  
GCTGTCGTTT CCAACTACCC GCTCAAAACGA TCTGTTTTTG CCAATTTATC ACAGCGCCGT 25020  
TGAAAAAGCC CTGCAAGATA AACTAGGCGT AAAGGCATTT AAGCTAAGCC CAACTAATGC 25080  
TCATACCGCT CGCGCGGCAA ATGAGAGCAG CCTAAATGCA GCCAATGGTG CCATTGCCCC 25140  
TAACAGCTCA AAAGTGGTGG CCGATGCACT TGGCCTTGGC GCGGCACAAC TAAGCCTAGA 25200  
TGCTGCCTGT GCTAGTTCGG TTTACTCATT AAAGCTTGCC TGGGATTACC TAAGCACTGG 25260  
CAAAAGCCGAT ATCATGCTAG CAGGCGCAGT ATCTGGCGCG GATCCTTTCT TTATTAAATAT 25320  
GGGATTCTCA ATCTTCCACG CCTACCCAGA CCATGGTATC TCAGTACCGT TTGATGCCAG 25380  
CAGTAAAGGT TTGTTTGCTG GCGAAGGCGC TGGCGTATTA GTGCTTAAAC GTCTTGAAGA 25440  
TGCCGAGCGC GACAATGACA AAATCTATGC GGTGTTAGC GCGGTAGGTC TATCAAAACGA 25500

FIG. 4A-25

31/134

CGGTAAAGGC CAGTTTGTAT TAAGCCCCTAA TCCAAAAGGT CAGGTGAAGG CCTTTGAACG 25560  
TGCTTATGCT GCCAGTGACA TTGAGCCAAA AGACATTGAA GTGATTGAGT GCCACGCAAC 25620  
AGGCACACCG CTTGGCGGATA AAATTGAGCT CACTTCAATG GAAACCTTCT TTGAAGACAA 25680  
GCTGCAAGGC ACCGATGCAC CGTTAATTGG CTCAGCTAAG TCTAACTTAG GCCACCTATT 25740  
AACTGCAGCG CATGCGGGGA TCATGAAGAT GATCTTCGCC ATGAAAAGAG GTTACCTGCC 25800  
GCCAAGTATC AATATTAGTG ATGCTATCGC TTCGCCGAAA AAACCTCTCG GTAAACCAAC 25860  
CCTGCCTAGC ATGGTTCAAG GCTGGCCAGA TAAGCCATCG AATAATCATT TTGGTGTAAG 25920  
AATCCGTCAC GCAGGCGTAT CGGTATTGG CTTTGGTGGC TGTAACGCCC ATCTGTTGCT 25980  
TGAGTCATAC AACGGCAAAG GAACAGTAAA GGCAGAAGCC ACTCAAGTAC CGCGTCAAGC 26040  
TGAGCCGCTA AAAGTGGTTG GCCTTGCCTC GCACTTTGGG CCTCTTAGCA GCATTAATGC 26100  
ACTCAACAAT GCTGTGACCC AAGATGGGAA TGGCTTTATC GAACTGCCGA AAAAGCGCTG 26160  
GAAAGGCCTT GAAAAGCACA GTGAACTGTT AGCTGAATT TTGGCTTAGCAT CTGCGCCAAA 26220  
AGGTGCTTAT GTGATAACT TCGAGCTGGA CTTTTTACGC TTTAAACTGC CGCCAAACGA 26280  
AGATGACCGT TTGATCTCAC AGCAGCTAAT GCTAATGCCA GTAACAGACG AAGCCATTCTG 26340  
TGATGCCAAG CTTGAGCCGG GGCAAAAAGT AGCTGTATTA GTGGCAATGG AAACCTGAGCT 26400  
TGAACGTGAT CAGTTCCGGG GCCGGGTTAA CTTGCATACT CAATTAGCGC AAAGTCTTGC 26460  
CGCCATGGGC GTGAGTTTAT CAACGGATGA ATACCAAGCG CTTGAAGCCA TCGCCATGGÀ 26520

FIG. 4A-26



32/134

CAGCGTGCTT GATGCTGCCA AGCTCAATCA GTACACCAGC TTTATTGGTA ATATTATGGC 26580  
GTCACGCGTG GCGTCACTAT GGGACTTTAA TGGCCCAGCC TTCACTATTT CAGCAGCAGA' 26640  
GCAATCTGTG AGCCGCTGTA TCGATGTGGC GCAAAACCTC ATCATGGAGG ATAACCTAGA 26700  
TGCGGTGGTG ATTGCAAGCGG TCGATCTCTC TGGTAGCTTT GAGCAAGTCA TTCTTAAAAA 26760  
TGCCATTGCA CCTGTAGCCA TTGAGCCAAA CCTCGAAGCA AGCCTTAATC CAACATCAGC 26820  
AAGCTGGAAT GTCGGTGAAG GTGCTGGCGC GGTCGTGCTT GTTAAAAATG AAGCTACATC 26880  
GGGCTGCTCA TACGGCCAAA TTGATGCACT TGGCTTTGCT AAAACTGCCG AAACAGCGTT 26940  
GGCTACCGAC AAGCTACTGA GCCAAACTGC CACAGACTTT AATAAGGTTA AAGTGATTGA 27000  
AACTATGGCA GGCCTTGCTA GCCAAATTCA ATTAGCGCCA ATAGTTAGCT CTCAAGTGAC 27060  
TCACACTGCT GCAGAGCAGC GTGTTGGTCA CTGCTTTGCT GCAGCGGGTA TGGCAAGCCT 27120  
ATTACACGGC TTAATTAACT TAAATACTGT AGCCCAAACC AATAAAGCCA ATTGCGCGCT 27180  
TATCAACAAT ATCAGTGAAA ACCAATTATC ACAGCTGTTG ATTAGCCAAA CAGCGAGCGA 27240  
ACAAACAAGCA TTAACCGCGC GTTTAAGCAA TGAGCTTAAA TCCGATGCTA AACACCAACT 27300  
GGTTAAGCAA GTCACCTTAG GTGGCCGTGA TATCTACCAG CATATTGTTG ATACACCGCT 27360  
TGCAAGCCTT GAAAGCATTG CTCAGAAATT GCGCAAGCG ACAGCATCGA CAGTGTCAA 27420  
CCAAGTTAAA CCTATTAAAG CCGCTGGCTC AGTCGAAATG GCTAACTCAT TCGAAACGGA 27480  
AAGCTCAGCA GAGCCACAAA TAACAATTGC AGCACAACAG ACTGCAAAACA TTGGCGTCAČ 27540

FIG. 4A-27

33/134

CGCTCAGGCA ACCAAACGTG AATTAGGTAC CCCACCAATG ACAACAAATA CCATTGCTAA 27600  
TACAGCAAAT AATTAGACA AGACTCTTGA GACTGTTGCT GGCAATACTG TTGCTAGCAA 27660  
GGTTGGCTCT GCGGACALAG TCAATTTTCA ACAGAAACCAA CAATTGGCTC AACRAAGCTCA 27720  
CCTCGCCTTT CTTGAAAGCC GCAGTGCGGG TATGAAGGTG GCTGATGCTT TATTGAAAGCA 27780  
ACAGCTAGCT CAAGTAACAG GCCAAACTAT CGATAATCAG GCCCTCGATA CTCAAAGCCGT 27840  
CGATACTCAA ACAAGCGAGA ATGTAGCGAT TGCCGCAGAA TCACCAGTTC AAGTTACAAC 27900  
ACCTGTTCAA GTTACAACAC CTGTTCAAAT CAGTGTTGTG GAGTTAAAAC CAGATCACGC 27960  
TAATGTGCCA CCATACACGC CGCCAGTGCC TGCAATTAAAG CCGTGTATCT GGAACATATGC 28020  
CGATTTAGTT GAGTACGCAG AAGGCGATAT CGCCAAGGTA TTTGGCAGTG ATTATGCCAT 28080  
TATCGACAGC TACTCGCGCC GCGTACGTCT ACCGACCACT GACTACCTGT TGGTATCGCG 28140  
CGTGACCAAA CTTGATGCGA CCATCAATCA ATTTAAGCCA TGCTCAATGA CCACTGAGTA 28200  
CGACATCCCT GTTGATGCGC CGTACTTAGT AGACGGACAA ATCCCTTGGG CGGTAGCAGT 28260  
AGAAATCAGGC CAATGTGACT TGATGCTTAT TAGCTATCTC GGTATCGACT TTGAGAACAA 28320  
AGGGGAGCGG GTTTATCGAC TACTCGATTG TACCCCTCACC TTCCTAGGCG ACTTGCCACG 28380  
TGGCGGAGAT ACCCTACGTT ACGACATTAA GATCAATAAC TATGCTCGCA ACGGCGACAC 28440  
CCTGCTGTTC TTCTTCTCGT ATGAGTGTTT TGTGGCGAC AAGATGATCC TCAAGATGGA 28500  
TGGCGGCTGC GCTGGCTTCT TCACTGATGA AGAGTTGCC GACGGTAAAG GCGTGATTCTG 28560

FIG. 4A-28

34/134

CACAGAAAGAA GAGATTAAAG CTCGCAGCCT AGTGCAAAAG CAACGCTTTA ATCCGTTACT 28620  
AGATTGTCCT AAAACCCCAAT TTAGTTATGG TGATATTTCAT AAGCTATTAA CTGCTGATAT' 28680  
TGAGGGTTGT TTTGGCCCAA GCCACAGTGG CGTCCACCAG CCGTCACTTT GTTTCGCATC 28740  
TGAAAAAATTC TTGATGATTG AACAAAGTCAG CAAGGTTGAT CGCACTGGCG GTACTTGGGG 28800  
ACTTGGCTTA ATTGAGGGTC ATAAGCAGCT TGAAGCAGAC CACTGGTACT TCCCATGTCA 28860  
TTTCAAGGGC GACCAAGTGA TGGCTGGCTC GCTAATGGCT GAAGGTTGTG GCCAGTTATT 28920  
GCAGTTCTAT ATGCTGCACC TTGGTATGCA TACCCAAACT AAAAATGGTC GTTTCCAACC 28980  
TCTTGAAAAAC GCCTCACAGC AAGTACGCTG TCGCGGTCAA GTGCTGCCAC AATCAGGCGT 29040  
GCTAACTTAC CGTATGGAAG TGACTGAAAT CGGTTTCAGT CCACGCCCAT ATGCTAAAGC 29100  
TAACATCGAT ATCTTGCTTA ATGGCAAGC GGTAGTGGAT TTCCAAAACC TAGGGGTGAT 29160  
GATAAAAGAG GAAGATGAGT GTACTCGTTA TCCACTTTTG ACTGAATCAA CAACGGCTAG 29220  
CACTGCACAA GTAAACGCTC AAACAAGTGC GAAAAAGGTA TACAAGCCAG CATCAGTCAA 29280  
TGCGCCATTA ATGGCACAAA TTCCTGATCT GACTAAAGAG CCAACAAGG GCGTTATTCC 29340  
GATTTCCCAT GTTGAAGCAC CAATTACGCC AGACTACCCG AACCGGTGAC CTGATACAGT 29400  
GCCATTACG CCGTATCACA TGTTTGAGTT TGCTACAGGC AATATCGAAA ACTGTTTCGG 29460  
GCCAGAGTTC TCAATCTATC GCGGCATGAT CCCACCACGT ACACCATGCG GTGACTTACA 29520  
AGTGACCACA CGTGTGATTG AAGTTAACGG TAAGCGTGGC GACTTTAAAA AGCCATCATC 29580

FIG. 4A-29

35/134

GTGTATCGCT GAATATGAAG TGCCCTGCAGA TGGCTGGTAT TTCGATAAAA ACAGCCACGG 29640  
CGCAGTGATG CCATATTCAA TTTTAATGGA GATCTCACTG CAACCTAACG GCTTTATCTC 29700  
AGGTTACATG GGCACAACCC TAGGCTTCCC TGGCCTTGAG CTGTTCTTCC GTAACCTAGA 29760  
CGGTAGCGGT GAGTTACTAC GTGAAGTAGA TTTACGTGGT AAAACCATCC GTAACGACTC 29820  
ACGTTTATTA TCAACAGTGA TGGCCGGCAC TAACATCATC CAAAGCTTTA GCTTCGAGCT 29880  
AAGCACTGAC GGTGAGCCTT TCTATCGCGG CACTGCGGTA TTTGGCTATT TTAAGGTGA 29940  
CGCACTTAAA GATCAGCTAG GCCTAGATAA CCGTAAAGTC ACTCAGCCAT GGCATGTAGC 30000  
TAACGGCGTT GCTGCAAGCA CTAAGGTGAA CCTGCTTGAT AAGAGCTGCC GTCACCTTAA 30060  
TGCGCCAGCT AACCAGCCAC ACTATCGTCT AGCCGGTGGT CAGCTGAACT TTATCGACAG 30120  
TGTTGAAATT GTTGATAATG GCGGCACCGA AGGTTTAGGT TACTTGATG CCGAGCGCAC 30180  
CATTGACCCA AGTGATTGGT TCTTCCAGTT CCACCTCCAC CAAGATCCGG TTATGCCAGG 30240  
CTCCTTAGGT GTTGAAGCAA TTATTGAAAC CATGCAAGCT TACGCTATTA GTAAAGACTT 30300  
GGGCGCAGAT TTCAAAAATC CTAAGTTTGG TCAGATTTTA TCGAACATCA AGTGGAAGTA 30360  
TCGCGGTCAA ATCAATCCGC TGAACAAGCA GATGCTATG GATGTCAGCA TTACTTCAAT 30420  
CAAAGATGAA GACGGTAAGA AAGTCATCAC AGGTAATGCC AGCTTGAGTA AAGATGGTCT 30480  
GCGCATATAC GAGGTCTTCG ATATAGCTAT CAGCATCGAA GAATCTGTAT AAATCGGAGT 30540  
GACTGTCTGG CTATTTTACT CAATTTCTGT GTCAAAAAGTG CTCACCTATA TTCATAGGCT 30600

FIG. 4A-30

36/134

GGCGCTTTT TTCTGGAAAT TGAGCAAAAG TATCTGCGTC CTAACTCGAT TTATAAGAAT 30660  
GGTTTAATTG AAAAGAACAA CAGCTAAGAG CCGCAAGCTC AATATAAATA ATTAAGGGTC 30720  
TTACAAATAA TGAATCCTAC AGCAACTAAC GAAATGCTTT CTCCGTGGCC ATGGGCTGTG 30780  
ACAGAGTCAA ATATCAGTTT TGACGTGCAA GTGATGGAAC AACAACTTAA AGATTTTAGC 30840  
CGGGCATGTT ACGTGGTCAA TCATGCCGAC CACGGCTTTG GTATTGGCA AACTGCCGAT 30900  
ATCGTGA CTG AACAAAGCGC AACAGCACA GATTACCTG TTAGTGCTTT TACTCCTGCA 30960  
TTAGGTACCG AAAGCCTAGG CGACAATAAT TTCCGCCGCG TTCACGGCGT TAAATACGCT 31020  
TATTACGCAG GCGCTATGGC AAACGGTATT TCATCTGAAG AGCTAGTGAT TGCCCTAGGT 31080  
CAAGCTGGCA TTTTGTGTGG TTCGTTTGA GCAGCCGGTC TTATTCCAAG TCGCGTTGAA 31140  
GCGGCAATTA ACCGTATTCA AGCAGCGCTG CCAAATGGCC CTTATATGTT TAACCTTATC 31200  
CATAGTCCTA GCGAGCCAGC ATTAGAGCGT GGCAGCGTAG AGCTATTTT AAAGCATAAG 31260  
GTACGCACCG TTGAAGCATC AGCTTTCTTA GGTCTAACAC CACAAATCGT CTATTACCGT 31320  
GCAGCAGGAT TGAGCCGAGA CGCACAAAGT AAAGTTGTGG TTGGTAACAA GGTATCGCT 31380  
AAAGTAAGTC GCACCGAAGT GGCTGAAAAG TTTATGATGC CAGCGCCCGC AAAAATGCTA 31440  
CAAAAACTAG TTGATGACGG TTCAATTACC GCTGAGCAA TGGAGCTGGC GCAACTTGTA 31500  
CCTATGGCTG ACGACATCAC TGCAGAGGCC GATTGAGTG GCCATACTGA TAACCGTCCA 31560  
TTAGTAACAT TGCTGCCAAC CATTTAGCG CTGAAAGAAG AAATTCAAGC TAAATACCAA 31620

FIG. 4A-31

37/134

TACGACACTC CTATTCGTGT CGGTGTGGT GCGGTGTGG GTACGCCCTGA TGCAGCGCTG 31680  
GCAACGTTTA ACATGGGCGC GCGTATATT GTTACCGGCT CTATCAACCA AGCTTGTGTT 31740  
GAAGCGGCG CAAGTGATCA CACTCGTAAA TTA CTGTTGCCA CCACTGAAAT GGCCGATGTG 31800  
ACTATGGCAC CAGCTGCAGA TATGTTTCGAG ATGGGCGTAA AACTGCAGGT GGTAAAGCGC 31860  
GGCAGCTAT TCCCAATGCG CGCTAACAG CTATATGAGA TCTACACCCG TTACGATTCA 31920  
ATCGAAGCGA TCCCATTAGA CGAGCGTGAA AAGCTTGAGA AACAAAGTATT CCGCTCAAGC 31980  
CTAGATGAAA TATGGGCAGG TACAGTGGCG CACTTTAAAG AGCGCGACCC TAAGCAAATC 32040  
GAACGCGCAG AGGTAACCC TAAGCGTAAA ATGGCATGTA TTTTCCGTTG GTACTTAGGT 32100  
CTTTCTAGTC GCTGGTCAA CTCAGGCGAA GTGGTCTGT AAATGGATTA TCAAATTTGG 32160  
GCTGGCCCTG CTCGCGTGC ATTTAACCA TGGCAAAAG GCAGTTACTT AGATAACTAT 32220  
CAAGACCGAA ATGCCGTCGA TTTGGCAAAG CACTTAATGT ACGGCGCGGC TTACTTAAAT 32280  
CGTATTAACT CGCTAACGGC TCAAGGCGTT AAAGTGCCAG CACAGTTACT TCGCTGGAAG 32340  
CCAAACCAA GAATGGCCTA ATACACTTAC AAAGCACCAG TCTAAAAAGC CACTAATCTT 32400  
GATTAGTGGC TTTTTTTTATT GTGGTCAATA TGAGGCTATT TAGCCTGTAA GCCTGAAAAAT 32460  
ATCAGCACTC TGACTTTTACA AGCAAATAT AATTAAGGCA GGGCTCTACT CATTATACT 32520  
GCTAGCAAAC AAGCAAAGTTG CCCAGTAAAA CAACAAGGTA CCTGATTAT ATCGTCATAA 32580  
AAGTTGGCTA GAGATTCGTT ATTGATCTTT ACTGATTAGA GTCGCTCTGT TTGGAAAAAG 32640

FIG. 4A-32

38/134

GTTCCTCGTT ATCATCAAAA TACACTCTCA AACCTTTAAT CAATTACAAC TTAGGCTTTC 32700  
TGCGGGCATT TTTATCTTAT TTGCCACAGC TGTATTTGCC TTTAGGTTTT GGGTGCAACT' 32760  
ACCATTAATT GAGGCCTCAT TAGTTAAATT ATCTGAGCAA GAGCTCACCT CTTTAAATTA 32820  
CGCTTTTCAG CAAATGAGAA AGCCACTACA AACCATTAAT TACGACTATG CGGTGTGGGA 32880  
CAGAACCTAC AGCTATATGA AATCAAACTC AGCGAGCGCT AAAAGGTACT ATGAAAAACA 32940  
TGAGTACCCA GATGATACGT TCAAGAGTTT AAAAGTCGAC GGAGTATTTA TATTCAACCG 33000  
TACAAAATCAG CCAGTTTTTA GTAAAGGTTT TAATCATAGA AATGATATAC CGCTGGTCTT 33060  
TGAATTAACT GACTTTAAAC AACATCCACA AAACATCGCA TTATCTCCAC AAACCAACA 33120  
GGCACACCCA CCGGCAAGTA AGCCGTTAGA CTCCCCCTGAT GATGTGCCTT CTACCCATGG 33180  
GGTTATCGCC ACACGATACG GTCCAGCAAT TTATAGCTCT ACCAGCATTT TAAAATCTGA 33240  
TCGTAGCGGC TCCCAACTTG GTTATTAGT CTTTCATTAGG TTAATTGATG AATGGTTCAT 33300  
CGCTGAGCTA TCGCAATACA CTGCCGCAGG TGTTGAAATC GCTATGGCTG ATGCCGCAGA 33360  
CGCACAAATTA GCGAGATTAG GCGCAAAACAC TAAGCTTAAAT AAAGTAACCG CTACATCCGA 33420  
ACGGTTAATA ACTAATGTCG ATGGTAAGCC TCTGTTGAAG TTAGTGCTTT ACCATACCAA 33480  
TAACCAACCG CCGCCGATGC TAGATTACAG TATAATAATT CTATTAGTTG AGATGTCATT 33540  
TTTACTGATC CTCGCTTATT TCCTTTACTC CTACTTCTTA GTCAGGCCAG TTAGAAAGCT 33600  
GGCTTCAGAT ATTAAAAAAA TGGATAAAG TCGTGAAATT AAAAAAGCTAA GGTATCACTA 33660

FIG. 4A-33

39/134

CCCTATTACT GAGCTAGTCA AAGTTGCGAC TCACCTTCAAC GCCCTAATGG GGACGATTCA 33720  
GGAACAAACT AACAGCTTA ATGAACAAAGT TTTTATTGAT AAATTAACCA ATATTCCCAA 33780  
TCGTGCGGCT TTTGAGCAGC GACTTGAAAC CTATTGCCAA CTGCTAGCCC GGCAACAAAT 33840  
TGGCTTTACT CTCATCATTG CCGATGTGGA TCATTTTAAA GAGTACAACG ATACTCTTGG 33900  
GCACCTTGCT GGGGATGAAG CATTAAATAA AGTGGCACAA ACACATATCG AACAGTTTAA 33960  
CCGTGCAGAA GATATTGTG CCCGTTTTGG TGGTGAAGAA TTTATTATGT TATTTGAGA 34020  
CATACCTGAT GAGCCCTTGC AGAGAAAGCT CGATGCGATG CTGCACCTCTT TTGCAGAGCT 34080  
CAACCTACCT CATCCAAACT CATCAACCGC TAATTACGTT ACTGTGAGCC TTGGGGTTTG 34140  
CACAGTTGTT GCTGTTGATG ATTTTGAATT TAAAAAGTGAG TCGCATATTA TTGGCAGTCA 34200  
GGCTGCATTA ATCGCAGATA AGGCGCTTTA TCATGCTAAA GCCTGTGGTC GTAACCAAGT 34260  
GTCAAAAACT ACTATTACTG TTGATGAGAT TGAGCAATTA GAAGCAATAA AAATCGGTCA 34320  
TCAAGCCTAA ACTCGTTTCA GTACTTTCCC CTAAAGTCAGA GCTATTTGCC ACTTCAAGAT 34380  
GTGGCTACAA GGCTTACTCT TTCAAAAACCT GCATCAATAG AACACAGCAA AATACAATAA 34440  
TTTAAGTCAA TTTAGCCTAT TAAACAGAGT TAATGACAGC TCATGGTGC AACTTATTAG 34500  
CTATTTCTAG CAATATAAAA ACTTATCCAT TAGTAGTAAC CAATAAAAAA ACTAATATAT 34560  
AAAACTATTT AATCATTATT TTACAGATGA TTAGCTACCA CCCACCTTAA GCTGGCTATA 34620  
TTCGCACTAG TAAAAATAAA CATTAGATCG GGTTACAGATC AATTTACGAG TCTCGTATAA 34680

FIG. 4A-34



40/134

AATGTACAAT AATTCACCTTA ATTTAATACT GCATATTTTT ACAAGTAGAG AGCGGTGATG 34740  
AAACAAAATA CGAAAGGCTT TACATTAATT GAATTAGTCA TCGTGATTAT TATTCTCGGT 34800  
ATACTTGCTG CTGTGGCACT GCCGAAATTC ATCAATGTTT AAGATGACGC TAGGATCTCT 34860  
GGGATGAGCG GTCAGTTTTC ATCATTTGAA AGTGCCGTAA AACTATACCA TAGCGGTTGG 34920  
TTAGCCAAAG GCTACAACAC TGCGGTTGAA AAGCTCTCAG GCTTTGGCCA AGGTAATGTT 34980  
GCATCAAGTG ACACAGGTTT TCCGTACTCA ACATCAGGCA CGAGTACTGA TGTGCATAAA 35040  
GCTTGTTGGT AACTATGGCA TGGCATTACC GATACAGACT TCACAATTGG TCGGTTAGT 35100  
GATGGCGATC TAATGACTGC AGATGTCGAT ATTGCTTACA CCTATCGTGG TGATATGTGT 35160  
ATCTATCGCG ATCTGTATTT TATTCAGCGC TCATTACCTA CTAAGGTGAT GAACTACAAA 35220  
TTTAAAACTG GTGAAATAGA AATTATTGAT GCTTTCTACA ACCCTGACGG CTCAACTGGT 35280  
CAATTACCAT AAATTTGGCG CTTATCTAAG TTGTACTTGC TCTGACCGAC ACAATAAATG 35340  
TCGTTTCTCA GCATATATCA AAATACACAG CAAAAATTG GGGTTAGCTA TATAGCTAAC 35400  
CCCAAAATCAT ATCTAACTTT ACACTGCATC TAATTCCAAA CAGTATCCAG CCAAAAGCCT 35460  
AAACTATTGT TGACTCAGCG CTAAAAATATG CGATGCAACA AACAAAGTCTT GGATCGCAAT 35520  
ACCTGAGCTA TCAAAAATGG TCACCTCATC AGCATTGTA CGTCCTGTTG CGGACTCGTT 35580  
TATCACCTGA CCAATCTCAA TTATCGGCGT ATTTCTGCTA TGTGAAACT CACCAATAAC 35640  
AATAGATTGA GAAGCAAAGT CGCAAAACAA GCGAGCATGA CTATATAGGT CAGTTGGCAA 35700

FIG. 4A-35

41/134

CTCTTGCTTA CCCACTTTAT CAGCGCCCAT TGCAGAAATA TGC GTTCCTG CTTGTACCCA 35760  
CTGCGCTTCA AATAAAGCG CTTGAGCTGT GGTGCTGTG ATAATAATAT CTGCTTGTTT 35820  
ACAAGCAGCT TGTGCATCAC AAGCTTCGGC ATTAATGCCT TTTTCTAATA AACGCTTAAC 35880  
CAAGTTTCA GTTTTGCTAG CACTACGGCC AACTACCAAT ACCTTAGTTA ATGAACGAAC 35940  
CTTGCTCACT GCTAGCACTT CATATTCAGC CTGATGACCG GTACCAAAAA CAGTTAATAC 36000  
CGTAGCATCT TCTCTCGCGA GGTAACCTAC TGCTACTGCA TCGGCAGCAC CAGTGCGGTA 36060  
AGCATTAAACG GTAGTGGCAG CAATCACCGN CTGCAACATA CCGGTTAATG GATCGAGTAA 36120  
AAATACGTTA GTGCCGTGGC ATGGTAAACC ATGTTTATGG TTATCAGGCC AATAGCTGCC 36180  
TGTTTTCCAG CCGACAAGGT TTGGCGTTGA AGCCGACTTT AATGAGAACA TTTTCATTAAG 36240  
GTTCCGCGCC TGTCGATTAA CTACCGGGAA CAAGTTGCT TTATCATCTA CGGCAGCGAC 36300  
AAACGCTTCT TTAACAGCGA TATAAGCCAG CTCATGGGAG ATGAGCTTTG ATGTTTGGC 36360  
TTCAGTTAAA TAGATCATAT TACCACCCCT GCACTCGATT CCAGATCTCA TAGCCACCAT 36420  
TATCACCATC AGTATCAAAT ACATGGTACT GAGCGTGCAAT TGAAGCTGTT GCACAGGCGT 36480  
GGTTCGGCAA AATATGTAGA CGACTACCTA CCGGGAACCTG CGCTAAAATCA ATACGCCGC 36540  
CATCAACTGC TTCAATAATG CCGTGCTCTT GATTAACAGT TATAACCTGT AGACCTGATA 36600  
ACACGTGACC GCTGTCGTCA CACACTAAAC CATAACCACA ATCTTTTGGC TGCTCTGCAG 36660  
TACCTCTATC ACCCGAAAGA GCCATCCAAC CCGCATCAAT GAAAATCCAG TTTTATCAG 36720

FIG. 4A-36

42/134

GATTATGACC AATAACACTG GTCACTACCG TTGCGGCAAT ATCAGTTAAC TGACACACGT 36780  
TTAGCCCTGC CATGACTAAA TCGAAGAAGG TGTACACACC CGCTCTAACC TCGGTGATCC 36840  
CATCAAGGTT TTGATAGCTT TCGCGTGTG GTGTTGAACC AATACTAACG ATGTCACATT 36900  
GCATACCCGC TCGCGGAATG CGTCAGCAGC TTGTACAGCC GCTGCAACTT CATTTTGGC 36960  
CGCATCAATT AATTGCTGTT TTTCAAAAACA TTGATATGAC TCACCAGCGT GAGTNAGTAC 37020  
GCCGTGAAAA CTCGCTGCGC CAGACGTTAG TATCTGAGCA ATTTCAATCA ACTTATCGGC 37080  
TTCCGGTGGA ATACCACCAC GATGGCCATC ACAATCAATT TCAATTAATG CTGGTATTG 37140  
GCAGTCATAA GAACCAACAGA AATGATTTAG CTGATGCGCT TGCTCAACAC TATCAAGTAA 37200  
AACTCTTGCA TTAATACCTT GGTCCAACAT TTTAGCAATA CGCGGCAACT TACCATCGGC 37260  
AATACCTACT GCATAAATAA TGTCTGTGA ACCTTTAGAT GCTAAGGCCT CGGCCTCTTT 37320  
TACCGTTGAT ACAGTGACTG GTGAGTTTTT AGTGGGTAAT AAAAACTCGG CTGCTTCAAG 37380  
TGATCTTAAC GTTTTAAAAAT GCGGTCCTAG GTTTGCACCT AATCCTTCAA TTTTTTGGCG 37440  
TAGTTGACTG AGGTTATTAA TAAATACTGG CTTATTTACA TATAAAAACG GTGTATCAAT 37500  
TGCTTGATAC TGACTTTGCT GAGTCGTGGA AAGTATTGA GTAGATGGCA TCTTTAATAT 37560  
CCTAGTTTCAAT CAATCAATCT AACAAAGTTG ATGCCTAGCC ACAGTGGCTT GTATTTCATGA 37620  
TGCTTTGGAA AATGCTTATA TTCAAAGTAT TTGAAAAGACA TCAAACTTCT TGTTTAATGC 37680  
TCAGTATCCA CCAGCACGCA TTTATTTTAT ATTAACCTATT ATCAAGATAT AGATTAGGTT 37740

**FIG. 4A-37**

CAAACCAAAT GATTAGTACT GAAGATCTAC GTTTATCAG CGTAATCGCC AGTCATCGCA 37800  
CCTTAGCTGA TGCCGCTAGA AACTAAATA TCACGCCACC ATCAGTGACA TTAAGGTTGC 37860  
AGCATATTGA AAAGAAACTA TCGATTAGCC TGATC 37895

**FIG. 4A-38**

44/134

6121

* MKQTLMAISI	MSLFSFNALA	AQHEHDHITV	DYEGKAATEH
TIAHNQAVAK	TLNFADTRAF	EQSSKNLVAK	FDKATADILR
AEFAFISDEI	PDSVNPSLYR	QAQLNMVPNG	YKVSDGIYQV
RGTDLSNLT	IRSDNGWIAY	DVLLTKEAAK	ASLQFALKNL
PKDGDVPVAM	IYSHSHADHF	GGARGVQEMF	PDVKVYGSDN
ITKEIVDENV	LAGNAMSRRRA	AYQYGATLGK	HDHGIVDAAL
GKGLSKGEIT	YVAPDYTLNS	EGKWETLTID	GLEMVFMDAS
GTEAESEMIT	YIPSKKALWT	AELTYQGMHN	IYTLRGAKVR
DALKWSKDIN	EMINAFGQDV	EVLFLASHSAP	VWGNQAINDF
LRLQORDNYGL	VHNQTLRLAN	DGVGIQDIGD	AIQDTIPESI
YKTWHTNGYH	GTYSHNAKAV	YNKYLGYFD	MNPANLNPLP
TKQESAKFVE	YMGGAADAIK	RAKDDYAQGE	YRFVATALNK
VVMAEPENDS	ARQLLADTYE	QLGYQAEGAG	WRNIYLTGAQ
ELRVGIQAGA	PKTASADVIS	EMDMPTLFDF	LAVKIDSQQA
AKHGLVKMNV	ITPDTKDILY	IELSNGNLSN	AVVDKEQAAD
ANLMVNKADV	NRILLGQVTL	KALLASGDAK	LTGDKTAFSK
IADSMVEFTP	DFEIVPTPVK		

\*  
8103

FIG. 4B

45/134

8186

\*  
STKASARVVA KFNVEEAAIS IQQCQGISLA FRYSDDLHGL  
LCHWDAANM QQEKAEILGL GSKQPEANPK NSSSELLALG  
IDQKLLVQRQ NLQHEVKHDA IADSIDVCHS LSKPANVGLF  
TESLASFDFA FSKLSLALGL GKAKIYSEKL AWLDFFRDRQ  
LAEPLALLAR KESESFYHSL ISHINTSNRC REIDVGFEIS  
ASDTEEKSAQ SAGKNDATCI GVLLWDGSHS VNFHVGTQAF  
QADSLRPKGK DGYEFRWENP RIESHQSLLA RLYGRVM

9016\*

**FIG. 4C**

8186

\*GCTAGTCTTA GCTGASRTHR YSAASRAGCT CGAACAACAG CTTTAAAATT  
CACTTCTTCT GCTGCAATAC TTATTTGCTG AACTGACCA ATACTCAGTG  
CAAAACGATA ACTATCATCA AGATGGAAAR GVAVAAAYSH ASNVAGGAAA  
ASRGNGNCYS GNGYSRAAHA RGTYSRASA SHSCCCAGTA AACAAATGCCA  
ATTATCAGCA GCGTTCATTT GCTGTTCTTT AGCCTCAATC AAACCTAAAC  
CAGACTTTTG TGGCTCAGCG TTAGGCTTAT TAGGYCYSHS TRASNASAAA  
AASNMTGNGN GYSAAGGYGY SRYSGNRGAA ASNRYSASNS RAACTCGACT  
CTAGTAAAGC AAGACCAATA TCTTGTTTTA ACAAACCTG TCGCTGATTA  
AGTTGATGCT CAACCTTGTG ATCCGCAATA GCATCGGAAA TSRSRGAAGY  
ASGNYSVAGN ARGGNASNGN HSGVAYSHSA SAAAAASSRA TCAACACAAT  
GGCTCAAGCT TTTAGGTGCA TTAACCTCAA GAAAAGTTTC GCTCAGTGCA  
GAGAAGTCAA ACGCAAAAGA TTTTAGCGAT AATGCCAGCA SVACYSHSSR  
SRYSRAAASN VAGYHTRGS RAASRHASHA AHSRYSSRAA CCAAGTCCTT  
TCGCTTTAAT GTAAGACTCC TTGAGCGCCC ACAAATCAA AAAGCGGTCT  
CGCTGCAAGG CCTCTGGTAA CGCTAACAAG GCTCGCTTTT GYGYSAAYS  
TYRSRGYSAA TRASHHARGA SARGGNAAGR AAAAARGYSG CTGATTCAGA  
GAAATAATGA CTAAGAATAG AGTGGATATT GGTGCTGTTA CGGCAACGCT  
CAATGTCGAC GCCAACTCA ATACTAGCAG AGTCAGTTTC SRGSRHTYRH  
SSRSRHSASN THRSRASNAR GCYSARGGAS VAGYHGSRAA SRASTHRGCT  
CCTTGCTTGC CTGACTGGCG CCTTTATTAT CAGCAGTGCA AATGCCTACT  
AATAGCCAAT CTCCACTATG ACTCACATTA AAGTGGACCC CGGTTTGAGY  
SSRAAGNSRA AGYYSASNAS AATHRCYSGY VATRASGYSR HSSRVAASNH  
HSVAGYTHRG NGCAAATTGC GCATCACTCA ATCTAGGCTT ACCTTTGTGCG

FIG. 4D-1

47/134

CCATATTCAA AGCGCCATTC ATTGGGGCGT ATTTCACTAT GTTGTGACAA  
TAAAGCGCGC AAAHGNAAS SRARGRYSGY YSASGYTYRG HARGTRGASN  
RARGGSRHSG NSRAAARGAA TAGCCTCTTA CCATTAAACC TTGAGTTTTA  
GCTTCTTGTT TAATGTAGCG ATTAACCTTA ATTAACCTCAT CTTCAGGCAG  
CCATGACTTA ACCAACTCTY RGYARGVAMT GYGNTHRYSA AGGNYSTYRA  
RGASNVAYSG ASGRTRSRYS VAGTGTAGTC TGGTTATCGC ACTCTTGTAT  
TGTTAACGGA CAGAAGTATA AGGAAATCAA<sup>\*</sup>

9157

**FIG. 4D-2**



9681

\*MSMFLNSKLS RSVKLAIASG LTASLAMPVF AEETAAEEQI ERVAVTGSRI  
AKAELTQPAP VVSLSAEELT KFGNQDLGSV LAELPAIGAT NTIIGNNNSN  
SSAGVSSADL RRLGANRTL VLVNGKRYVAG QPGSAEVDLS TIPTSMISRV  
EIVTGGASAI YGSDAVSGVI NVILKEDFEG FEFNARTSGS TESVGTQEH  
FDILGGANVA DGRGNVTFYA GYERTKEVMA TDIRQFDWAG TIKNEADGGE  
DDGIPDRLRV PRVYSEMINA TGVINAFGGG IGRSTFDSNG NPAAQQRD  
TNSFAFGSFP NGCDTCFNT AYENYIPGVE RINVGSSFN DFTDNIQFY  
DFRYVKSDIQ QQFQPSFRFG NININVEDNA FLNDDLQRM LDAGQTNASF  
AKFFDELGNR SAENKRELFR YVGGFKGGFD ISETIFDYDL YVYGETNNR  
RKTLDLIPD NFVAAVDSVI DPDTGLAACR SQVASAQGD YDTPASVNGS  
DCVAYNPFGM GQASAEARDW VSADVTREDK ITQQVIGGTL GTDSEELFEL  
QGGAIAMVVG FEYREETSGS TTDEFTKAGF LTSAATPDSY GEYDVTEYFV  
EVNIPVLKEL PFAHELSDG AYRNADYSHA GKTEAWKAGM FYSPLQLAL  
RGTVGEAVRA PNIAEAFSPR SPGFGRVSDP CDADNINDDP DRVSNCAALG  
IPPGFQANDN VSVDTLSGGN PDLKPETSTS FTGGLVWTPT FADNLSFTVD  
YYDIQIEDAI LSVATQTVAD NCV DSTGGPD TDFCSQVDRN PTYDIELVR  
SGYLNAAALN TKGIEFQAAY SLDLESFNAP GELRFNLLGN QLELERLEF  
QNRPEINDE KGEVGDPELQ FRLGIDYRLD DLSVSWNTRY IDSVVTDVS  
ENGGSPEDLY PGHIGSMTH DLSATYYINE NFMINGGVRN LFDALPPGYT  
NDALYDLVGR RAFLGIKVM

\*  
12590

FIG. 4E

13040

\*MAKINSEHLD EATITSNKCT QTETEARHRN ATTTPEMRRF IQESDLSVSQ  
LSKILNISEA TVRKWRKRDS VENCNTPHH LNTTLTPLQE YVVVGLRYQL  
KMPLDRLLKA TQEFINPNVS RSGLARCLKR YGVSRSVDIQ SPHVPMRYFN  
QIPVTQGS DV QTYTLHYETL AKTLALPSTD GDNVVQVVSL TIPPKLTEEA  
PSSILLGIDP HSDWIYLDIY QDGNTQATNR YMAYVLKHGP FHLRKLLVRN  
YHTFLQRFPG ATQNRPSKD MPETINKTPE TQAPSGDS<sup>\*</sup>  
13903

**FIG. 4F**

50/134

13906

\* MSQTSKPTNS ATEQAQDSQA DSRLNKRLKD MPIAIVGMAS IFANSRYLKN  
FWDLISEKID AITELPSTHW QPEEYYDADK TAADKSYCKR GGFLPDVDFN  
PMEFGLPPNI LELTDSSQLL SLIVAKEVLA DANLPENYDR DKIGITLGVG  
GGQKISHSLT ARLOYPVLKK VFANSGISDT DSEMLIKKFQ DQYVHWEENS  
FPGSLGNVIA GRIANRFDG GMNCVDAAC AGSLAAMRMA LTELTEGRSE  
MMITGGVCTD NSPSMYMSFS KTPAFTTNET IQPFDIDSKG MMIGEGIGMV  
ALKRLEDAER DGDRIYSVIK GVGASSDGKF KSIYAPRPSG QAKALNRAYD  
DAGFAPHTLG LIEAHGTGTA AGDAAEFAGL CSVFAEGNDT KQHIALGSVK  
SQIGHTKSTA GTAGLIKAAL ALHHKVLPPPT INVSQSPKPL DIENSPFYLN  
TETRPWLPRV DGTPRRAGIS SFGFGGTNFH FVLEEYNQEH SRTDSEKAKY  
RQRQVAQSFL VSASDKASLI NELNVLAASA SQAEFILKDA AANYGVRELD  
KNAPRIGLVA NTAEELAGLI KQALAKLAAS DDNAWQLPGG TSYRAAAVEG  
KVAALFAGQG SQYLNMGRLD TCYYPEMRQQ FVTADKVFAA NDKTPLSOTL  
YPKPVFNKDE LKAQEAILTN TANAQSAIGA ISMGQYDLFT AAGFNADMVA  
GHSFGELSAL CAAGVISADD YYKLAFARGE AMATKAPAKD GVEADAGAMF  
AIITKSAADL ETVEATIAKF DGVKVANYNA PTQSVIAGPT ATTADAAL  
TELGKAINL PVSGAFHTEL VGHAQAPFAK AIDAAKFTKT SRALYSNATG  
GLYESTAABI KASFKKHMLQ SVRFTSQLEA MYNDGARVVFV EFGPKNILQK  
LVQGTLVNTE NEVCTISINP NPKVDSDLQL KQAAMQLAVT GVVLSIEDPY  
QADIAAPAKK SPMSISLNAA NHISKATRAK MAKSLTGIV TSQIEHVIEE  
KIVEVEKLVE VEKIVEKVVE VEKVVEVEAP VNSVQANAIQ TRSVVAPVIE  
NQVVSKNKSP AVQSISGDAL SNFFAAQQQT AQLHQQFLAI PQQYGETFTT  
LMTEQAKLAS SGVAIPESLQ RSMEQFHQLQ AQTLOSHTQF LEMQAGSNIA  
ALNLLNSSQA TYAPAIHNEA IQSQVVQSQT AVQPVISTQV NHVSEQPTQA  
PAPKAQPAPV TTAVQTAPAQ VVRQAAPVQA AIEPINTSVA TTTPSAFSAE

FIG. 4G-1

51/134

TALSATKVQA TMLEVVAEKT GYPTEMLELE MDMEADLGID SIKRVEILGT  
VQDELPGLE LSPEDLAECR TLGEIVDYM SKLPAEGSMN SQLSTGSAAA  
TPAANGLSAE KVQATMMSVV AEKTYGPTM LELEMDMEAD LGIDSIKRV  
ILGTVQDELP GLPELSPEDL AECRTLGEIV DYMNSKLADG SKLPAEGSMN  
SQLSTSAAAA TPAANGLSAE KVQATMMSVV AEKTYGPTM LELEMDMEAD  
LGIDSIKRV ILGTVQDELP GLPELNPEDL AECRTLGEIV TYMNSKLADG  
SKLPAEGSMH YQLSTSTAAA TPVANGLSAE KVQATMMSVV ADKTYGPTM  
LELEMDMEAD LGIDSIKRV ILGTVQDELP GLPELNPEDL AECRTLGEIV  
DYMNSKLPAE GSANTSAAAS LNVSAVAPQ AAATPVSNGL SAEKVQSTMM  
SVVAEKTGYP TEMLELGMDM EADLGIDSIK RVEILGTVQD ELPGLEPNP  
EDLAECRTLGEIVDYMNSKL ADGSKLPAEG SANTSATAAT PAVNGLSADK  
VQATMMSVVA EKTGYPTM ELGMDMEAD LGIDSIKRV ILGTVQDELP  
LPELNPEDL AECRTLGEIVS YMNSQLADGS KLSTSAAEGS ADTSAANA  
PAAISAEPV ELPPHSEVAL KKLNAANKLE NCFAADASV INDDGHNAGV  
LAEKLIKQGL KVAVVRLPKG QPQSPLSSDV ASFELASSQE SELEASITAV  
IAQIETQVGA IGGFIHLQPE ANTEEQTAVN LDAQSFTVHS NAFLWAKLLQ  
PKLVAGADAR RCFVTVSRID GGGFYLNTDA LKDAELNQAA LAGLTKTL  
EWPQVFCRAL DIATDVDATH LADAITSELF DSQAQLPEVG LSLIDGKVN  
VTLVAAEAAD KTAKAELNST DKILVTGGAK GVTFEALAL ASRSQSHFIL  
AGRSELQALP SWAEGKQTSE LKSAAIAHII STGQKPTPKQ VEAAVWPVQS  
SIEINAALAA FNKVGASA EYVSM DVTSAA ITAALNGRSN EITGLIHGAG  
VLADKHIQDK TLAEKLVYK TKVNLKALL AALEPSKIKL LAMFSSAAGF  
YGNIGQSDYA MSNDILNKAA LQFTARNPQA KVMSFNWGPW DGGMVNPALK

FIG. 4G-2

52/134

KMFTERGVYV IPLKAGAELE ATQLLAETGV QLLIGTSMQG GSDTKATETA  
SVKKLNAGEV LSASHPRAGA QKTPLQAVTA TRLLTPSAMV FIEDHRIGGN  
SVLPTVCAID WMREAASDML GAQVKVLDYK LLKGIVFETD EPQELTLELT  
PDDSDEATLQ ALISCNGRPQ YKATLISDNA DIKQLNKQFD LSAKAITTA  
ELYSNGTLFH GPRLQGIQSV VQFDDQGLIA KVALPKVELS DCGEFLPQTH  
MGGSQPFAED LLLQAMLVWA RLKTGSASLP SSIGEFTSYQ PMAFGETGTI  
ELEVIKHNR SLEANVALYR DNGELSAMFK SAKITISKSL NSAFLPAVLA  
NDSEAN

<sup>\*</sup>  
22173

**FIG. 4G-3**

53/134

22203

\*MPLRIALILL PTPQFEVNSV DQSVLASYQT LQPELNALLN SAPTPEMLSI  
TISDDSDANS FESQLNAATN AINNGYIVKL ATATHALLML PALKAAQMRI  
HPHAQLAAMQ QAKSTPMSQV SGELKLGANA LSLAQTNALS HALSQAKRNL  
TDVSVNECFE NLKSEQQFTE VYSLIQQLAS RTHVRKEVNQ GVELGPKQAK  
SHYWFSEFHQ NRVAAINFIN GQQATSYVLT QGSGLLAACS MLNQQRMLFI  
LPGNSQQQIT ASITQLMQQL ERLQVTEVNE LSLECQLELL SIMYDNLVNA  
DKLTTRDSKP AYQAVIQASS VSAAKQELSA LNDALTALFA EQTNATSTNK  
GLIQYKTPAG SYLTLTPLGS NNDNAQAGLA FVYPGVGTVY ADMLNELHQY  
FPALYAKLER EGDLCAMLQA EDIYHLDPKH AAQMSLGDLA IAGVGSSYLL  
TQLLTDEFNI KPNFALGYSM GEASMWASLG VWQNPHALIS KTQTDPLFTS  
AISGKLTAVR QAWQLDDTAA EIQWNSFVVR SEAAPIEALL KDYPHAYLAI  
IQGDTCVIAG CEIQCKALLA ALGKRGIAAN RVTAMHTQPA MQEHQNVMDF  
YLQPLKAELP SEISFISAAD LTAKQTVSEQ ALSSQVVAQS IADTFCQTLN  
FTALVHHAQH QGAKLFVEIG ADRQNCTLID KIVKQDGASS VQHQPCTVP  
MNAKGSQDIT SVIKALGQLI SHQVPLSVQP FIDGLKRELT LCQLTSQQLA  
AHANVDSKFE SNQDHLLQGE V

24515

FIG. 4H

54/134

24518

\*MSLPDNASNH LSNQKGASQ ASKTSKQSKI AIVGLATLYP DAKTPQEFWQ  
NLLDKRDSRS TLTNEKLGAN SQDYQGVQGG SDRFYCNKGG YIENFSFNAA  
GYKLPEQSLN GLDDSFLLWAL DTSRNALIDA GIDINGADLS RAGVVMGALS  
FPTTRSNDLF LPIYHSAVEK ALQDKLGVKA FKLSPTNAHT ARAANESSLN  
AANGAIAHNS SKVVADALGL GGAQLSLDAA CASSVYSLKL ACDYLSTGKA  
DIMLAGAVSG ADPFFINMGF SIFHAYPDHG ISVPFDASSK GLFAGEGAGV  
LVLKRLEDAE RDNDKIYAVV SGVGLSNDGK GQFVLSNPVK GQVKAFERAY  
AASDIEPKDI EVIECHATGT PLGDKIELTS METFFEDKLQ GTDAPLIGSA  
KSNLGHLLTA AHAGIMKMIF AMKEGYLPPS INISDAIASP KKLF GKPTLP  
SMVQGWPDKP SNNHFGVRTR HAGVSFVGFG GCNAHLLLES YNGKGTVKA  
ATQVPRQAEP LKVVGSLASHF GPLSSINALN NAVTQDGNGF IELPKKRWKG  
LEKHSSELLAE FGLASAPKGA YVDNFELDFL RFKLPPNEDD RLISQQMLML  
RVTDEAIRDA KLEPGQKVAV LVAMETELEL HQFRGRVNLH TQLAQSLAAM  
GVSLSTDEYQ ALEAIAMDSV LDAAKLNQYT SFIGNIMASR VASLWDFNGP  
AFTISAAEQS VSRCIDVAQN LIMEDNLDAV VIAAVDLSGS FEQVILKNAI  
APVAIEPNLE ASLNPTSASW NVGEGAGAVV LVKNEATSGC SYGQIDALGF  
AKTAETALAT DKLLSQATD FNKVKVIETM AAPASQIQLA PIVSSQVTHT  
AAEQRVGHCF AAAGMASLLH GLLNLNTVAQ TNKANCALIN NISENQLSQL  
LISQTASEQQ ALTARLSNEL KSDAKHQLVK QVTLGGRDIY QHIVDTPLAS  
LESITQKLAQ ATASTVVNQV KPIKAAGSVE MANSFETESS AEPQITIAAQ  
QTANIGVTAQ ATKRELGTPP MTTNTIANTA NNLDKTLETV AGNTVASKVG  
SGDIVNFQQN QQLAQQAHLA FLESRSAGMK VADALLKQQL AQVTGQTIDN  
QALDTQAVDT QTSENVIAIA ESPVQVTTTPV QVTTTPVQISV VELKPDHANV  
PPYTPPVPAL KPCIWNYADL VEYAEGDIAK VFGSDYAIID SYSRRVRLPT  
TDYLLVSRVT KLDATINQFK PCSMTTEYDI PVDAPYLV DG QIPWAVAVES  
GQCDLMLISY LGIDFENKGE RVYRLDCTL TFLGDLPRGG DTLRYDIKIN  
NYARNGDTLL FFFSYECFVG DKMILKMDGG CAGFFTDEEL ADGKGVIRTE

FIG. 4I-1

55/134

EEIKARSLVQ KQRFNPLLDC PKTQFSYGDI HKLLTADIEG CFGPSHSGVH  
QPSLCFASEK FLMIEQVSKV DRTGGTWGLG LIEGHKQLEA DHWYFPCHFK  
GDQVMAGSLM AEGCGQLLQF YMLHLGMHTQ TKNGRFQPLE NASQQVRCRG  
QVLPQSGVLT YRMEVTEIGF SPRPYAKANI DILLNGKÄVV DFQNLGVMIK  
EEDECTRYPL LTESTTASTA QVNAQTSAKK VYKPASVNAP LMAQIPDLTK  
EPNKGVIPIS HVEAPITPDY PNRVPDTPVF TPYHMFEFAT GNIENCFGPE  
FSIYRGMIPP RTPCGDLQVT TRVIEVNGKR GDFKKPSSCI AEYEVPADAW  
YFDKNSHGAV MPYSILMEIS LQPNGFISGY MGTTLGFPGL ELFFRNLDGS  
GELLREVDLR GKTIRNDSRL LSTVMAGTNI IQSFSFELST DGEPPFYRGTA  
VFGYFKGDAL KDQLGLDNGK VTQPWHVANG VAASTKVNLL DKSCRHFNAP  
ANQPHYRLAG GQLNFIDSVE IVDNGGTEGL GYLYAERTID PSDWFFQFHF  
HQDPVMPGSL GVEAIIETMQ AYAISKDLGA DFKNPKFGQI LSNIKWKYRG  
QINPLNKQMS MDVSITSIKD EDGKKVITGN ASLSKDGLRI YEVFDIAISI  
EESV

\*  
30529

**FIG. 4I-2**



56/134

30730

\*  
MNPTATNEML SPWPWAVTES NISFDVQVME QQLKDFSRA  
YVNHADHGF GIAQTADIVT EQAANSTDLP VSAFTPALGT  
ESLGDNNFRR VHGVKYAYYA GAMANGISSE ELVIALGQAG  
ILCGSFGAAG LIPSRVEAAI NRIQAALPNG PYMFNLIHSP  
SEPALERGSV ELFLKHKVRT VEASAFGLGT PQIVYYRAAG  
LSRDAQGKVV VGNKVIKVS RTEVAEKFMM PPAKMLQKL  
VDDGSITAEQ MELAQLVPMA DDITAEADSG GHTDNRPLVT  
LLPTILALKE EIQAKYQYDT PIRVGCGGGV GTPDAALATF  
NMGAAIYIVTG SINQACVEAG ASDHTRKLLA TTEMADVMTA  
PAADMFMGV KLQVVKRGTL FPMRANKLYE IYTRYDSIEA  
IPLDEREKLE KQVFRSSLDE IWAGTVAHFN ERDPKQIERA  
EGNPKRKMAL IFRWYLGSLSS RWSNSGEVGR EMDYQIWAGP  
ALGAFNQWAK GSYLDNYQDR NAVDLAKHLM YGAAYLNRIN  
SLTAQGVKVP AQLLRWKPNQ RMA

\*  
32358

FIG. 4J

57/134

32834

\*  
MRKPLQTINY DYAVWDR TYS YMKSNSASAK RYYEKHEY PD  
DTFKSLKVDG VFIFNRTNQP VFSKGFNHRN DIPLVFELTD  
FKQHPQONIAL SPQTKQAHPP ASKPLDSPDD VPSTHGVIAT  
RYGPAIYYSS TSILKSDRSG SQLGYLVFIR LIDEWFIAEL  
SQYTAAGVEI AMADAADAQL ARLGANTKLN KVTATSERLI  
TNVDGKPLLK LVLYHTNNQP PPMLDYSIII LLVEMSFLLI  
LAYFLYSYFL VRPVRKLASD IKKMDKSREI KKLRYHYPIT  
ELVKVATHFN ALMGTIQEQT KQLNEQVFID KLTNIPNRRRA  
FEQRLETYCQ LLARQQIGFT LIIADVDFHK EYNDTLGHLA  
GDEALIKVAQ TLSQQFYRAE DICARFGGEE FIMLFRDIPD  
EPLQRKLDAM LHSFAELNLP HPNSSTANYV TVSLGVCTVV  
AVDDFEFKSE SHIIGSQAAL IADKALYHAK ACGRNQALSK  
TTITVDEIEQ LEANKIGHQ

\*  
34327

**FIG. 4K**

1  
\*  
AATAGATCGACTCGCAAAAGTTGCTTAAGATAGTGTCAATATAGCTTCTTATTTGTA  
AATATTGTTTTTTATGTGTAAACATGTTTAGTGTGTGTAAATGCTGTTAATTATCCT  
TTTGGGATTGTAATAGCTGATGTTGCTGGCTAATGAGTACTTTTAGTTCGGCAATAT  
CTTGCTTTAAATCGCTAACTTCAGTTTTTAATTCACCCACACTTGTTGTATTTTTAA  
GGCTCTCTTCCCCACCATCGACAAACCAGGATGATATGAAACCGGTAAACGTACCAA  
AGAGACCGACACCTGCAGTCATGAGTAATGCCGCAATGATACGTCCGCCAGTGGTGA  
CGGGGTAGTAGTCACCGTAACCAACAGTCGTTATTGTCACAAATGACCACCAAAGTG  
CGTCGATGCCGTTATTGATGTTACTGCCTACTTGATCCTGTTCTAACAATAAAATAC  
CGATAGCACCAAAGGTGACAAGGATGAAGGATATCGCAGATACCAGCGAAAAGGTGG  
CTTTAAACCGATGTTCAAAAATCATTTTTTAAGATAATTTTTTGATGAGCGTATATTCT  
GAATAGATCTTAATACTCTAGCGATACGAATTATGCGAATAAACTGCAGTTGCTCGA  
CCATCGGAATACTCGACAGTAGGTCAATCCAACCCCATTTTCATAAACTGAAATTTAT  
TCTCAGCTTGGTGAAAGCGAATTACAAAGTCAGTGAAAAAGAATAAGCAAATCGTAT  
TATCTACGCTCGTTAATATTTTCAGTGACGTTACTTGAAAAGGTAAAAATAAGTTGCA  
GTAGTGATGATACGACCACATGAAGTGATAAAATAAGCATGAAAATCTGAAATGGAT  
TTACATCACTGTTGTTTTTGGTGCCACTTTTAAGGTTTCGTTTTTCACAACTCTGCTGCC  
TCGGTTCATTGATTTTTGTTAATATAAACCTTAGTCAGTAGCAAGACAAAATATATTT  
ACATCAATGTCATCGTATTATTCAACCGCGCGTCGTGTATTCAGACCAAGATCGTTG  
TATATGTTAGTCATGTAGCGATGAGATTATCATGCGACAGGAGAGAATTATGTTTGT  
TATTATTTTTTACGTACCTAAAGTTAATGTTGAAGAAGTAAAACAGGCGTTATTTAA  
CGTCGGAGCTGGCACCATCGGTGATTATGATAGTTGTGCTTGGCAATGTTTGGGGAC  
TGGGCAGTTCCAACCTTTACTTGGTAGCCAGCCACATATTGGTAAGCTAAATGAGGT  
TGAATTCGTTGATGAGTTTAGAGTAGAAATGGTTTGTGAGCAGAAAATGTAAGGGC  
AGCAATAAATGCACTTATTGCTGCGCACCCTTATGAAGAACCTGCTTATCATATTCT  
GCAAACATTGAATCTTGATGAGTTACCTTAAGTTAGATGCACTGCACTTAATTGGTT  
CGCTGTGCTAGGTTAGCAATTAGCAATTTTGACCATGTTAGCGATAGTTTTGGCACA

FIG. 5-1

59/134

AGTGATCGATATTAACTATCCGATTCAGATCCCATTTTACTGCTGAATTAGGTTT  
CATTACACTTGTTCTAGTGGTTTTTCCCGACAGGTGTAACCTCTGTTACTTGCGTAAG  
GTTGATAATCTCTACCGCATTGGCAGGAGTTACACCTGCACCAGGCATAATACTAAT  
TCTACCATCTGCTTGGTTAACTAACGTTTGGATTAAGGCGCAGCCTTCTAGCGCTTG  
AGCTTGTTGACCAGAGGTTAAAATACGCTCACAACCAGCAGTGATCAAGGTCTCCAA  
GGCTTGTTGTGGATCATTACACAAGTCGAAAGCGCGGTGGAAGGTTACGCCGAGATC  
ACGTGATGCCACCATTAAAGCGTTTTAAAGCTGGCTCGTCAATATTACCATCTGCTGT  
TAACGCGCCAATAACGACCCCTTGGACACCGAGTAACTTCATGAATTTGATGTCGGA  
AACCATAATATCAACTTCTTGTTTCGCTATATACAAAATCACCGGCGCGAGGGCGAAT  
AATGGCATAAATGGGGATCGTTGCTAGATCAATAGACTTTTGTACAAAACCTGCGTT  
GGCGGTCAAGCCACCTAATGCTAATGCCGAGCACAACTCAATACGATCGGCGCCAGA  
TGCTTGAGCCGTCAGCAGTGATTCTATATTATCGACACATACTTCTATTGTCATTGT  
CATATACTTCTCTTTAAAAAGTTTATTAAAAATAATAAAGCCAGCATAAGTCGTTTT  
ATACAATATGAAAGGGGAAAAGGCGACTTAGCTCGCCTAGATCAATTATTATGGCAG  
AATACTGCCGTATTGTGATTAGAAAGACAGTTTTTTAAGCTCAATAGCCGTTATCGC  
GTTGTTATCTACCATCGTGTAACTTTTCTGGCCTGGGTGCTTTATTAACTGTTTC  
AGTGGCTGGATTAGGGTGAAATGATTCTTTTTTCAAATCTGTTTTTTTGTATTTGAA  
CGTACCTGTAATGTCTTGCTGCTCACGAAGACGTACAAATATTGGTTGCGCATAGCT  
TGGTAGTGCCGCATTGACATGTTGATAGAATTCAGACGCTGAAAATTCATGAATAGG  
GCAATTCAAAGTCAGCGCGACCATGCCTGCTCGGCCATCGTGATGTGGGAGCTTGAC  
ACCATAAGCCACACTTTGCTCAATTTGCACAAAATCGTTAACTTGAGCTTCTACTTG  
CGTCGTGGCGACATTTTCACCTTTCCAGCGGAATGTATCACCTAATCTATCCACAAA  
GGAAATATGGCGATAACCTTGGTAATGAACGAGATCGCCGGTATTAAAATAACAGTC  
ACCGTCTTTTAATACTGACTTAAATAGCTTTTTATTACTTTTCGTTGTCATCGGTATA  
ACCATCAAATGGTGAACGTTTAGTTATCTTTGTTAGCAGTAGCCCTGTTTCTCCCGT

**FIG. 5-2**

60/134

TTTTACTTTGGTCATTTTCCCTTTTCGCATTATACACAGGTTTGTCAATTGTCAATATC  
ATATTGTATGACGGTAAAAGCAAGTGGAGTAACCCCGCTGTATGCGGTAAGTTCAG  
CGCATTGGAGAACACAAGATTACACTCACTGGCGCCATAGAATTCATTAATATGCTC  
GATCCCAAACGTTGTTGGAAATGATCCCAAATTTCGGGGCGTAATCCATTACCTAT  
GATTTTCTTTATATTATGCTGTTTGTCTTTATTGCTAGGCGGTACATTTAATAAATA  
ACGGCAGAGCTCGCCGATGTAAGTAAACGCAGTGGCATTATGAGCACGAACTTCATC  
CCAAAAGCGACTTGAACTGAATTTTTTCAGAAAGTGCGAGGGTTGCTGCGCTACCAA  
CACGGCGCTTAATGACACTGTCAGTGCATTGTTATGGTATAGGGGGAGTGATAAATA  
CAATACATCATCAGCTGTTAAGCGTAATGATGCCATCCCCATGCCTGCCATGGATTT  
AAACCAACGGTGATGGCTCATTCTTGCTGCTTTTGGCAGTCCAGTTTTTCCCGAGGT  
AAAGATATAAAACGCGCAATGCTTAAGCTGTATTTGTGCTGTTGATTACAGGGTTCAA  
TACTGAATATCCTGCGACTAGTGTAGATATGTTTTTATAACCATCACTCATGTCTGG  
CGTTTCTAAAGCGGGTACGTAAAAGACATTCTGTTGTAATGTGATGACAAATTGGT  
TTCAATATTATTAATGGCGGATGTGTATAGTTCATCTGCGATGAGTAATTTGGTATC  
GACCACGCTAAGACTATGTTGAGGATTGAATCCCGTTGTGTCGTATTTATCATACA  
AGCAATCGCGCCAAGCTTGACAACTGCGAGGGCAATAATGATGGTTTCAGGCCTGTT  
ATCGAGCATGATGGCGACTTTATCATTTTTACCAATGCCGTATTCATGAAGGAAATG  
GGCATATTGATTTGCTTGCTTATTCAATGAATCGTAACTATAACGCTGGTCTTTAAA  
TTGTATTGCGATCAAGTCAGAGTTATTGACAGCTTGCTGCTCTAGTAATAAACCAAT  
AGACATAAAACGTTTCGGGCTTTGCTTGTTGTAAGTGCCATAAGCCTTTGATGATTGG  
CTTTGGGGTTTTTAATAGATTGATGGTACTTTTCAGGAATTGTTTGCCGGTTATAAC  
AGTCATAAGCTAATTCTTTTTATCAAGAAGAGGGGTTATGACACCAAATAAATGGGT  
CACGCGTTGGTTTAATTTGGTTAGACTAAATGTGTTGTTTTGCTGTGATAATGCGAC  
GTTCAAACAACTTGAGAAGGTAAAAAATAGCATTTTTTAAATTGAACATCAATACT  
AATGTGTTGAATATCAATCAAGTTTTCTAACTGTGCGAGCACGCGTGCTTTAGCAAA

FIG. 5-3

61/134

CATGCCATGTGCTATTGCTGTTTTAAACCCCATTAGTTTCGCTGGGATAAAATGTAA  
ATGGATTGGATTTGTGTCTTTGGAGATATAAGCATATTTATATACGTCAAAGGACT  
AAATTTAAACAATGAAATCGGCTCGTAAGCATAATTCGCTGGCGTATTTACTATTTT  
CTCACCGCTGGAACGTTGAGATCGTTGGCACGTTTTTCGCTGTTTTCGTTTTCTGTAA  
GAATGTCGATGTACACTCCCACGCAAATTGTCCATCTACAAACACATCAATATGAGT  
ATCAATGAAACGTCCTGTATCCGTTATGTACTCCTTAATTACACGACATGTGCTCGT  
CAATATCGCGTTTTAATGCTATCGGTTGATGTTGTGTTATGCGATTTTCGATAATGGAC  
TAGTCCTAATATAGATATCGGAAATTGTGTTGATGTCATGAGTTTCATCAATAATGG  
AAAGATCATCACAATGGATAAGTAACCGGTACATAGTTTGTGTTATTAAACCCACA  
GCATTTAATATATTGCTTTAAATTTTCGCTGATCTATTTTTTGTCCACTGATACTAAA  
TTGCTCAGTACACACTTGTGTGCGACCAAGTGTTTCATCAGTGTTTTAACAATTGTATT  
GAECTACTGCTTTACATATAAAAGCGAGATAATCGGTTGCTTTGTTAACAGTGTTGAT  
CTGGTTAGCGTGCAATTGAAATAATTCATATAAGAGTATGTAGCATTTATGTTAATAT  
TTTGTTTTGGAAGTTGAATTGGCGAATCCGTAATCGGTTTATGGCAGTTCGGTCAAA  
TACTTCAGGTAACTCGTTACTCATACCATTGATAGTGTTAAAGTGATTGACTGAAT  
AAAGAATAGAGCTAAAAGTGGAATAATTATGCAAGATGCGGGTATGTTATTACGCAT  
TGCTTATGAGGCAATGAAAGAGTTAGAGGTTGATGTCATTGAAGTACTTTCTCGTTG  
TAACATAAGTGAAGAAGTACTGAATGATAAGGATCTTCGCACACCTAATCATGCACA  
AACACATTTTTGGCAAGTATTAGAAGACATATCACAAGATCCTAACATCGGCATTTTC  
ACTTGGTGAGAGAATGCCAGTGTTACGGGGCAGGTATTACAGTATCTTTTTCTCAG  
TAGTCCTACATTTGGTACTGGCTGGGAACGCGCAACAAAATACTTTTCGATTAATCAG  
TGATGCGGCGAGTGTTTCTATCAAGATGGAAGGCTGTGAAGCGCGATTATCTGTGAA  
CTTAGATGGTTTAGCGGAAGATGCGAATCGTCATTTGAATGATTGCCTAGTGATCGG  
TGCATTTAAATTTTGTTTATATGTGACAGAAGGCGAATTTAAAGTAAGCAAAATAGC  
CTTTGCTCATGCTCGCCCGAAAGATATTACTGCCTATACCAATGTATTTACATGTCC

**FIG. 5-4**

62/134

GATTGAGTTTGCTGCCGAAGATAATTATATTTATTTTCGATGCTGATTTACTCGAACG  
TCCTTCTTCGCATGCGGAGCCTGAGCTATTCGCCTTACACGATCAGCTTGCAAGCCG  
TAAAATAGCCAAGTTAGAACTGCAAGATTTAGTGGATAAAGTACGTAAGGTTATTGC  
ACAACAACCTTGAGTCTGGTGTGGTGACTTTAGAAAGTATCGCCACTGAACTTGACAT  
GAAACCACGTATGCTAAGAGCGAAGTTAGCTGACATTGATTATAACTTTAATCAAAT  
ACTCGCTGATTTTCGTTGCGAGTTATCAAAAAAAGTGTGGCGAATACGGACGAGTC  
TATTGATCAGATTGTCTATCTCACTGGTTTTTCTGAACCAAGTACTTTTTATCGTGC  
CTTTAAGCGCTGGGTAAAATGACGCCAATTGAATATCGCCGTAGCAAAGTTCGCGGT  
TAGGCATGCTAATCAACACGAGTCCTAAAAATTGCTGCTTAGTGATAGTGCATAG  
TGCATAGTGCTAGTAAGCCAAGTACAAAGCGTTAAAGTTAAGTACTTGAGCGAACCA  
TCAGACACCACTTACTAGATTAAGCACCTATTAATGATTGACCACAAATTCTGATCG  
TATTGCCTGTGATCCCTGCAGCTTGAGGTTGCGCAAAAAAGCTATCGCTTCAGCAA  
CATCAACTGGCTTACCACCTTGTTTTAATGAATTCATACGACGACCAGCTTCACGAA  
CTGTAAATGGAATCGCTGCTGTCATTTTTGTTTTCAATAAAGCCTGGTGCAACAGCAT  
TAATGGTGATGTATTTGTCTGCAAGCGGAGTTTGCATTGCATCAACATAACCAATGA  
CTGCGGCCTTAGACGTTGCATAATTAGTCTGACCAAAGTTACCCGCAATCCCACTCA  
TCGAAGACACACAAACAATGCGGCCATAGTCGTTGAGCAGATCATCATTTAGCAGTC  
GCTCATTGATTCTTTCCATTGCCGACAAGTTAATATCCATCAGTACATCCCAATGGT  
TATCCGGCATACTGCTAGCGTTTTGTCTTTTGTACCCCGGCATTATGGACGATGA  
TATCAAGCGACTGTTCTCGCACAAAGTCAGCAATGATATTTGGGGCGTCAGCAGCGG  
TAATATCAGCAACAATGCTGCTACCTTTCAAGCAATGAGCTACTTTTTCAAGGTCCT  
GTTTTAATGCCGGAATGTCTAAGCAAATAACATGTGCGCCATCACGGGCGAGTGTTT  
CAGCAATAGCAGCCCCGATGCCACGTGATGCACCAGTGACAAGTGCTGTCTTTCCTT  
GTAATGGTTTTGCCGTGTTACTTGTTTCGTTAATAACTTCGTTAATAACTTCGTTAA

**FIG. 5-5**

63/134

TAACCTTCGTTAATAGCCCCATTAATCGAACCGGGTTTTACGTTAATAACCTGTGCTG  
AGATATAGGCTGATTTTGCTGAGGTTAAGAAACGTAGCGGGGCCTCTAATAATTGCT  
CACTACCAGGTTGTACATAGATAAGTTGACAGGTA CTACCATTCTTGCCTATTTCTT  
TGGCGACACTGCGACAAAACCCTTCTAAAGATCTTTGTACAGTCGCGTAGCTTACAT  
CGTCAAGATGTTCACTCGGATGACCTAACACGATCACTCTGCTGCATGGCGAGAGCT  
GCTTAATTACAGGTTGAAAAAACGATGTAATGCACTTAATTGCTTGCTGTTCTTAA  
TGCCTGAGGCGTCGAAGATAATACCGTTGAAGCGATCTGTTTTAGCGATAGCATTAA  
GGCTAATAGGTGTCGCGACTAAAGACGTTTGATTAAATTCAATATTAAGATCGGCTA  
ACGCTGACGTGTTATTAGGATAAGAAATCGTGACTTCAGCATCTTTAAATGTGTTAA  
GAATGGGTTTAATTAATTTGCTGTTGCTGGCTGCGCCGATGAGTAAGTTGCCAGAGA  
TGAGATCGGTTCCCTGATCGTAGCGTGTTAACGTAACCGGTCGTGGCAGATTAAGCG  
CTTTAAATAAACCTGATGTCCACTTGCCATTAGCGAGTTTTGCGTATGTATCCGTCA  
TTTTCTAATCCTTGTTATAGTGAACAGTTTGAATCTCGAAGATGTACATGTGTTAA  
AATTATCTGATAGCTATGACTTATCTGCCACTACGTAATAATAAATAGACCAGTTCA  
TTACATCGTTAATCGATATAGTATAACTAAATACTAAGTAAATTATAATGATAAGAC  
TGTTATCGTACTCGGATCAAACCTCTGATCAGCAAATAATCAAATTAGAGTTTTTATT  
TTAAACTTGTATCAACAATGTTACATTAATGTATCTTACGTCTAATGTGCTACGGGC  
ATATTTAAGTCACTAAATTAAAGGAATAAACCATGACAGGTCAAACAATAAGAAGAG  
TAGCAATTATCGGCGGTAACCGTATCCCGTTTGACGTTCAAATACAGCGTATTCAA  
AACTAAGTAACCAAGATATGCTGACGGAACTATCCGTGGCTTGGTGGTTAAATATA  
ACCTACGTGGTGAACAACCTGGGGGAAGTTGTTGCTGGTGCGGTAATTAAGCATTCTC  
GTGATTTTAACTTAACACGTGAAGCCGTGCTAAGTGCAGGTCTTGACCTGAAACGC  
CTTGTTATGACATTCAACAAGCTTGTGGTACTGGTCTAGCTGCAGCTATCCAAGTAG  
CAAACAAAATTGCGCTTGGTCAAATAGAAGCGGTATTGCTGGTGGTTCTGATACGA

FIG. 5-6



64/134

CATCAGATGCACCGATTGCAGTCAGTGAAGGCATGCGTAGTGTATTACTTGAGCTTA  
ATCGAGCTAAAACGGGTAAGCAACGTTTGAAAGCACTATCTCGTCTACGTCTAAAAC  
ACTTTGCGCCACTAACGCCTGCAAATAAAGAGCCGCGTACCAAATGGCGATGGGCG  
ATCATTGTCAAGTAACAGCGAAAGAGTGGAATATCTCACGTGAAGCACAAAGATGCAT  
TGGCCTGCGCAAGTCATCAAAAATTAGCTGCAGCATATGAAGAAGGTTTCTTTGATA  
CGTTAGTTTCACCTATGGCCGGCTTAACGAAAGATAACGTATTACGCGCAGATACAA  
CAGTTGAGAACTGGCTAAATTGAAACCTTGTTTTGATAAAGTAAACGGCACTATGA  
CGGCGGGTAACAGTACTAACCTTACCGATGGAGCATCAGCTGTATTACTTGCAAGTG  
AAGAATGGGCAGCGGCACATAACTTACCAGTACAAGCTTATCTAACATTTGGTGAAA  
CGGCCGCTATCGACTTCGTTGATAAGAAAGAAGGTCTGTTAATGGCGCCTGCATACG  
CAGTGCCAAAAATGTTGAAGCGTGCTGGCCTTACATTACAAGACTTCGATTACTATG  
AAATACATGAAGCATTGTGCTGCGCAGTTATTAGCAACGCTAGCAGCTTGGGAAGACG  
AAAAATTCTGTAAAGAAAACTGGGTCTAGATGCTGCGCTTGTTCAATTGATATGA  
CCAAGTTAAACGTGAAAGGGAGTAGCTTAGCCACGGGTACCCATTTGCCGCAACTG  
GTGGTCGTGTTGTGCTACGCTAGCGCAATTACTTGATCAGAAAGGTTCAAGTCTGTG  
GTTTGATCTCGATTTGTGCTGCTGGTGGTCAAGGTATCACGGCAATTTTAGAGAAAT  
AAACGCACTGTTTATTATCTATTGATTAAGCTGTCCTGAGATACTGGATATTTTTAA  
ATAAAACGCCAATACTGCAGAGTATTGGCGTTTTTTTGTAATACCAATTCCTATATA  
ACGGTGCAATTTTAAACACTTAATTTCCGGCATTGGTATCATAAAAAAGCAGCACCGA  
AGTGCTGCTTGATTGTAGATTAACCTATTAAAATAGAGAGGCTAGAATTAGTCTTCG  
TATGCTTCATTATGTACGCCAGCTGCACGACCCGATGGATCAGCATTGTTTTTGAAA  
CTTTCATCCCAAGCTAATGCTTCTACAGTTGAACAAGCAACGGATTTACCAAACGGT  
ACGCATTTTCGCTGCTGAATCACCTGGGAAGTGATCTTCAAAGATGGCACGATAGTAG  
TAACCTTCTTTTCGTATCTGGTGTGTTAATTGGGAACCTAAATGCTGCACTTGCTAAC  
ATTTGATCAGTTACCGCTTCTTCAACGTGTACTTTAAGTTGGTCAATCCAAGAATAA

FIG. 5-7

65/134

CCAACACCATCAGAGAATTGTTCTTTTTGACGCCATACAATTTCTTCAGGTAGTAAA  
TCTTCAAATGCTTCTCGAATGATGTTTTCTCAATGCGGTCGCCC GTGATCATTTTT  
AGTTCAGGGTTTAGACGCATTGACGCATCAACAAATTCTTTATCTAAGAAAGGAACA  
CGTGCTTCGATGCCCCAAGCTGCCATAGATTTGTTTGACGTAAGCAATCAAACATA  
TGTAATTTATTTACTTTACGTACCGTCTCTTCATGGAATTCTTTTCGCATTTGGCGCT  
TTGTGGAAGTACAAGTAACCACCGAACAGTTCATCAGCACCTTCACCAGAAAGCACC  
ATCTTAATCCCCATGGCTTTAATTTTACGTGCCATTAGGTACATAGGGGTTGATGCA  
CGAATTGTTGTTACATCGTAGGTTTCAATGTGGTAAATCACGTGCGGTAAAGCGTCG  
ATACCTTCTTGACAGTAAATTCAATTGAATGATGGATAGTACCTAAGTGATCTGCC  
ACTTTTTGTGCAGCGGCTAAATCTGGAGAACCATTTAGGCCTACAGAGAAAGAGTGT  
AGTTGTGGCCACCATGCTTCGGTTTTACCACCGTCTTCAATACGACGTTTTGCATAC  
TGTTGGGTGATTGCTGAAATAACAGATGAATCTAACCCGCCTGATAATAATACGCCG  
TAAGGTACATCACACATTAATTGACGTTTAACTGCATCTTCCAAACCTTGCTTAACA  
ACGCTTTTATCACCACCATTTTGTGCAACGTTATCAAATCTTTCCAATCACGTTGA  
TAATAAGGCGTGACTACACCATCCTTACTCCACAGGTAATGACCTGCTGGGAATTCT  
TCAATTTGAGTACAAATTGGCACTAGTGCTTTTCATTTT CAGAGGCAACATAAAAGTTA  
CCGTGTTTCATCATAGCCCGTATAAAGAGGGATGATACCGATATGGTCACGGCCAATC  
AGGTAAGCGTCCTCTGTTTCGT CATATAAAGCGAAAGCAAAAATACCATTTAGATCA  
TCTAAAAATTGTGTGCCTTTTTCTTTATATAGCGCAAGTATCACTTCGCAATCTGAT  
TCTGTTTGGAATTCAAAGTCTACGTT CAGCGTTTTCTTTAAATCTTGTGGTTATAA  
ATTTACCATTAAACAGCAAGTACGTGTGTCTTTCTTCATTATATAGCGGCTGTGCA  
CCATTATTTACATCGACAATAGCAAGACGTT CATGAACTAAAATAGCATTGTCACTT  
GTATAGATACCTGACCAATCTGGGCCGCGGTGACGTAGTAACTTTGATAGTTCTAGT  
GCTTGTTTCGCGAAGAGGTTTAATGTCTGATTTGATGTCTAGAATTCCGAATATTGAG

**FIG. 5-8**

66/134

CACATAACTAATTCCTTCTGGGGCTGCGTCTGCAGCTAACTTTCTAAATAGTGTGTC  
TAATTTGCCACATTGTAGATTTAATGCAACATTAATGATAAAAACATTTATAAAAAA  
TGTAATTCAATGTGGAATCGATAATTTAATGGCTTAAAAGTGAAGATCCATTAATTG  
TGATGGCGAGGTGATAGACCAATGTAGACCTTAATGAATAAAGCAGGCACGATTGAA  
TCCATTCAACGCAAAGTGGTACTAACTATTGTTTTAAACGTTATAAATAGTGTTTTA  
AAGGTTATAAGTAAATAATTTAAAAACAATAATAATCCACATGCATTAAATTTATCA  
TGATAAACCGCTATATCTCAATGGCAATTTGGGATAAGTGTAATAATATATGTAAAT  
GAATGAGTTGACTTGCTTTTTTTTACACTAAGTGATGAAATTAAAGCTAGATGTCGTT  
GTTAGCATTGATTAATAACGTACTAAAATACGACATCTAGTATAGAAATTTAAAAAA  
CAGTTGGTTTTGATAGCATAACTGCATAAACTAATCAGCTTATTGTCTGTAATATTT  
TTGTAATTTAAATAGGTTTAATAAAATTATATGTCTGATAAATATAAACCGTACGAC  
CTTTCCTTTAAAAAGACGTTTTTGTCTGCCTAAGTTTTGGCCTGTGTGGTTCGGGGTG  
TTTGCAATATACTTATTAGCTTTTATGCCAGTAAAGCCGCGTGATAAATTTGCTCGA  
TTCATAGCGAAGAAATTGTTTAGTCTAAAAATGATGGCAAAGCGTAAAAAGGTAGCA  
AAGATCAATTTATCTATGTGCTTCCCTGAAATGGATGATACGGAACAAGACCGTATA  
ATCATGGTCAATCTAGTTACTTTTTGTCAAACCTATCTTAAGTTATGCAGAGCCAAGT  
GCGCGTAGTCGTGCTTATAACCGTGACCGTATGATAGTGCATGGTGGCGAGAATTTA  
TTTCCGCTACTTGAACAAGGTAAGGCTTGATCTTATTAGTGCCGCATAGCTTCGCT  
ATTGATTTTGCAGGTTTACACATTGCTTCTTATGGCGCGCCATTTTGTACTATGTTT  
AACAAATTCTGAGAATGAGTTGTTGATTGGCTGATGACACGTCAACGCGCTATGTTT  
GGAGGCACTGTTTATCACCGCAAGGCAGGGCTAGGGGCTCTAGTTAAATCACTTAAG  
AGCGGTGAAAGCTGTTATTACTTACCTGATGAAGACCATGGACCTAAGCGTAGTGTA  
TTTGCGCCTTTATTTGCGACTCAAAAAGCAACTTTACCTGTAATGGGCAAGCTAGCA  
GAAAAAACAAATGCACTCGTTGTTCTGTTTATGCGGCATATAATGAATCACTAGGT  
AAATTTGAAACCTTTATTTCGACCAGCAATGCAAACTTTCCATCAGAAAGCCCAGAA  
CAAGATGCAGTGATGATGAATAAAGAGATTGAAGCCTTGATTGAATGTGGTGTGAT

FIG. 5-9

67/134

CAATATATGTGGACACTTAGATTATTGAGAACACGTCCGGACGGTAAAAAATCTAC  
TAATAAAGTTTAATAAACACCATAATCTTCGTTGAATATGGTGTTTACCCCCCTGAA  
TACCCCTCTAAATTAATAACAAAAAAGCCATTTACGTAACATCTAATGATGATTTAG  
CCTGCACTTGCTTTGTTTTAGTCTTAAGAGCCTAATAAACTTGATCTAGGTATAGA  
TTCTGTCCTTTCTTTACGTAACGCGATCTATTTTTTTTAACCGATAGTTGTTATAATT  
AGTTTCATATGAAAGAGATATCGTTTCAGTAAAAGCTATTTTCGTTTCAATAGATAAT  
TTATTTATAGTCATATTTTCTGTAATGACAATCATTTTCTCATCTAGACTATAGATA  
AGAATACGAATTAAGTAAGAACATTAATTTTACAAGAATATAAAATATCCCATCGGA  
GCTATAAGAATGAAAAAGACTAAAATTGTTTGTACAATTGGTCCAAAACTGAATCA  
GTAGAGAACTAACAGAGCTTGTTAATGCAGGCATGAACGTTATGCGTTTAAATTTC  
TCTCATGGTAACTTTGCTGAACATTCAGTGCGTATTCAAATATCCGTCAAGTAAGT  
GAAACCTGAATAAGAAAATTGCTGTTTTACTGGATACTAAAGGTCCAGAAATCCGT  
ACGATTAACTAGAAAACGGTGACGATGTAATGTTGACCGCTGGTCAGTCATTACG  
TTTACAACAGACATTAACGTGGTAGGTAATAAAGACTGTGTTGCTGTAACATATGCT  
GGTTTTGCTAAAGACCTTAATCCTGGTGCAATCATCCTTGTTGATGATGGTTTAATT  
GAAATGGAAGTTGTTGCAACAACCTGACACTGAAGTTAAATGTACAGTATTAAATACT  
GGTGCACTTGGTGAAAATAAAGGCGTTAACTTACCTAACATCAGTGTAGGTCTACCT  
GCATTGTCAGAAAAAGATAAAGCTGATTTAGCGTTTGTTGTGAGCAAGAAGTTGAT  
TTTGTTGCTGCATCATTTATTTCGTAAGGCTGATGATGTAAGAGAAATTCGTGAAATC  
CTATTTAATAATGGTGCGGAAAACATTCAGATTATCTCGAAAATTGAAAACCAAGAA  
GGTGTAGACAATTTTCGATGAAATCTTAGCTGAATCAGACGGTATCATGGTTGCTCGT  
GGCGATCTCGGTGTTGAGATCCCAGTTGAAGAAGTGATCATGGCACAGAAGATGATG  
ATCAAAAAATGTAATAAAGCAGGTAAAGTTGTAATTACTGCAACACAAATGCTTGAT  
TCAATGATCAGTAACCCACGTCCAACACGTGCAGAAGCGGGCGATGTTGCCAATGCT  
GTGCTTGACGGTACCGACGCGGTAATGCTTTCTGGTGAACTGCGAAAGGTAAATAC

FIG. 5-10

CCAGTTGAAGCTGTGTCTATCATGGCAAACATCTGTGAACGTACTGATAACTCAATG  
TCTTCGGATTTAGGTGCGAACATTGTTGCTAAAAGCATGCGCATTACAGAAGCTGTG  
TGTAAGGTGCGGTAGAAACAACAGAAAAATTGTGTGCTCCACTTATTGTTGTTGCA  
ACTCGTGGCGGTAAATCAGCAAAATCTGTTCTGTAAATACTTCCCGAAAGCAAATATT  
CTTGCTATCACAACAAATGAAAAAGCAGCGCAACAGTTATGCCTAACTAAAGGCGTA  
AGCAGCTGCATCGTTGAGCAGATTGATAGCACTGATGAGTTCTACCGTAAAGGTAAA  
GAGCTTGCAATTAGCAACTGGTTTAGCTAAAGAAGGCGATATCGTTGTTATGGTATCA  
GGTGC GTTAGTACCATCAGGTACAACGAATACGGCATCTGTTACCAACTTTAAGTT  
GCCATATTGATATTATAAAAAAGAGAGCGTATGCTCTCTTTTTTTATATCTGTAGTT  
TATATGTCTGTACAAAAAATGATAAAGAGTACATAAACTATTAATATAGCGTAATA  
TATAATGATTAACGGTGATGAAAGGGTTAAATAAATGGATAGTGCTAAACATAAAAT  
TGGCTTAGTCCTTTCTGGCGGTGGTGCGAAAGGTATTGCTCATCTTGGTGTATTAAA  
ATACCTGTTAGAGCAAGATATAAGACCGAATGTAATTGCGGGTACAAGTGCTGGCTC  
TATGGTTGGTGCACTTTATTGCTCAGGACTTGAGATTGATGACATTTTACAATTCTT  
CATCGATGTAAAACCTTTTTCTTGGAAGTTTACCCGTGCCCCGTGCTGGCTTTATAGA  
CCCCGCAAAATTATATCCTGAAGTGCTAAATATATCCCCGAGGATAGCTTTGAGTA  
CCTTCAACCTGAATTGCGCATTGTTGCCACCAACATGTTACTCGGTAAAGAGCATAT  
ATTTAAAGATGGCTCCGTGATTAATGCCTTATTAGCATCAGCCAGCTACCCTTTAGT  
TTTTTCTCCGATGATCATTGACGATCAAGTGTATTGAGATGGCGGTATTGTTAATCA  
TTTCCCCGTGAGTGTCATTGAAGATGATTGCGATAAAATAATCGGCGTATACGTGTC  
GCCCATTTCGTGAGGTGCAAGCTGACGAACTCTCGAGTATAAAAGACGTGGTATTACG  
TGCGTTCACGCTGCAGGGTAGTGGTGCTGAATTAGATAAACTATCGCAATGTGATGT  
GCAAAATTTATCCAGAAGCGCTATTGAATTACAATACGTTTGCAACCGATGAAAAATC  
ATTACGGGAGATCTACCAGATTGGTTATGATGCTGCAAAAGATCAACATGACAACCT  
TATGGCATTGAAAGAAAGTATCACCACCAGCGAGGTTAAAAAGAACGTCTTTAGCAA

**FIG. 5-11**

69/134

ATGGTTTGGTGATAAACTTGCTAGCAACAGCGGCAAATAGCGGCCACACGGATTTA  
TACACTAGGATAATGGGCGTTAATAGCCTCACTGTCGTTGTGTGGTCTCTAATTTTA  
GCTAAATCTTGTGTTATACTGACTTCCTATTAATCATAAACGATTTATCACGGTAAA  
CATGACTCAAATAAATAACCCGCTTCACGGCATGACACTCGAAAAAGTAATTAACAG  
TCTCGTTGAACAATATGGCTGGGATGGTCTTGGATACTACATCAACATTCGTTGCTT  
TACTGAAAATCCAAGTGTTAAGTCTAGTCTTAAATTTTTACGTAAAACCCCTTGGGC  
ACGTGATAAAGTAGAAGCGCTATATATCAAATGGTGACTGAAGGCTAACTGTCTCC  
ACGCTAGCGAACCGCTGTTTATAGTTAATATAAGTACTATAAGCAGGGCTCGTTAAT  
TCAGTATGTAATTAATCCTGAATACCTCCGCTTATTTCAACATTGTACTCTCTAGAT  
AACACTCTCAACATTACACCTTCAACATCACAGCCTCCACATAACATCCGATGACAT  
AGCCCTGTTATTTTTTACATTTATCTATATGCTATATATTTTAGCCATTTGATCAAT  
TGAGTTAATTTCTGCAATGACAAAGATATAACCATCATCCAGTACAAATTTATTATGA  
AGATACCGACCATTCTGGTGTTGTTTACCACCCTAACTTTTTTAAATACTTTGAACG  
TGCACGTGAGCATGTGATAAATAGTGACTTACTAGCAACATTGTGGAATGAACGCGG  
TTTAGGTTTTGCGGTGTATAAAGCCAATATGACTTTTCAGGATGGGGTCTGAATTTGC  
TGAAGTGTGTGATATTCGCACTTCTTTTGTCTAGACGGTAAGTACAAAACGATCTG  
GCGCCAAGAAGTATGGCGTCCGAATGCGACTAGGGCTGCCGTTATCGGTGATATTGA  
AATGGTGTGCTTAGACAAACAAAACGTTTACAGCCCATCCCTGATGATGTGTTAGC  
TGCAATGGTTAGTGAATAAATGGTTCATGCATAAATAGTTAATACATGATTCTGGCC  
CGTCACGTTTACAGATAAGAGGCATCCGATGCCTCCTTCTTATTACCAATACTACTG  
CTTATCCCTTTCTAACTATCTTTAGCGTCCATAACACACTGAGCATTTATTCTATTA  
ATCAGTGATTGTGATTTAATTATCTTCTATATATGTAATTTAATGTAATTTTCAATT  
TATTTTGTAGCTACATTAAGGCTTACGAATGTACGCTAAAATGAGATGTCAGACTAAT  
TTTAGCTTATTAATCTGTTAGCCGTTTATATTTTATAAAGATGGGATTTAACTTAAA

**FIG. 5-12**

70/134

TGCAATTAATTATGGCGTAAATAGAGTGAAAACATGGCTAATATTCAC TAAGTCCTG  
AATTTTATATAAAGTTTAATCTGTTATTTTAGCGTTTACCTGGTCTTATCAGTGAGG  
TTTATAGCCATTATTAGTGGGATTGAAGTGATTTTAAAGCTATGTATATTATTGCA  
AATATAAATTGTAACAATTAAGACTTTGGACACTTGAGTTCAATTTCGAATTGATTG  
GCATAAAATTTAAAACAGCTAAATCTACCTCAATCATTTTAGCAAATGTATGCAGGT  
AGATTTTTTTTCGCCATTTAAGAGTACACTTGTACGCTAGGTTTTTGT TTAGTGTGCA  
AATGAACGTTTTTGATGAGCATTGTTTTTAGAGCACAAAATAGATCCTTACAGGAGCA  
ATAACGCAATGGCTAAAAAGAACACCACATCGATTAAGCACGCCAAGGATGTGT TAA  
GTAGTGATGATCAACAGTTAAATTCTCGCTTGCAAGAATGTCCGATTGCCATCATTG  
GTATGGCATCGGTTTTTGCAGATGCTAAAACTTGGATCAATTCTGGGATAACATCG  
TTGACTCTGTGGACGCTATTATTGATGTGCCTAGCGATCGCTGGAACATTGACGACC  
ATTACTCGGCTGATAAAAAAGCAGCTGACAAGACATACTGCAAACGCGGTGGTTTCA  
TTCCAGAGCTTGATTTTGATCCGATGGAGTTTGGTTTACCGCCAAATATCCTCGAGT  
TAACTGACATCGCTCAATTGTTGTCATTAATTGTTGCTCGTGATGTATTAAGTGATG  
CTGGCATTGGTAGTGATTATGACCATGATAAAATTGGTATCACGCTGGGTGT CGGTG  
GTGGTCAGAAACAAATTTTCGCCATTAACGTGCGGCCTACAAGGCCCGGTATTAGAAA  
AAGTATTAAGCCTCAGGCATTGATGAAGATGATCGCGCTATGATCATCGACAAAT  
TTAAAAAAGCCTACATCGGCTGGGAAGAGAACTCATTCCCAGGCATGCTAGGTAACG  
TTATTGCTGGTCGTATCGCCAATCGTTTTGATTTTGGTGGTACTAACTGTGTGGTTG  
ATGCGGCATGCGCTGGCTCCCTTG CAGCTGTAAAATGGCGATCTCAGACTTACTTG  
AATATCGTTTCAGAAGTCATGATATCGGGTGGTGTATGTTGTGATAACTCGCCATTCA  
TGTATATGTCATTCTCGAAAACACCAGCATTTACCACCAATGATGATATCCGTCCGT  
TTGATGACGATTCAAAGGCATGCTGGTTGGTGAAGGTATTGGCATGATGGCGTTTA  
AACGTCTTGAAGATGCTGAACGTGACGGCGACAAAATTTATTCTGTACTGAAAGGTA  
TCGGTACATCTTCAGATGGTCGTTTTCAAATCTATTTACGCTCCACGCCCAGATGGCC  
AAGCAAAAGCGCTAAAACGTGCTTATGAAGATGCCGGTTTTGCCCTGAAACATGTG

FIG. 5-13

71/134

GTCTAATTGAAGGCCATGGTACGGGTACCAAAGCGGGTGATGCCGCAGAATTTGCTG  
GCTTGACCAAACACTTTGGCGCCGCCAGTGATGAAAAGCAATATATCGCCTTAGGCT  
CAGTTAAATCGCAAATTGGTCATACTAAATCTGCGGCTGGCTCTGCGGGTATGATTA  
AGGCGGCATTAGCGCTGCATCATAAAATCTTACCTGCAACGATCCATATCGATAAAC  
CAAGTGAAGCCTTGATATCAAAAACAGCCCGTTATACCTAAACAGCGAAACGCGTC  
CTTGATGCCACGTGAAGATGGTATTCCACGTCTGTCAGGTATCAGCTCATTTGGTT  
TTGGCGGCACCAACTTCCATATTATTTTAGAAGAGTATCGCCCAGGTCACGATAGCG  
CATATCGCTTAAACTCAGTGAGCCAACTGTGTTGATCTCGGCAAACGACCAACAAG  
GTATTGTTGCTGAGTTAAATAACTGGCGTACTAAACTGGCTGTCGATGCTGATCATC  
AAGGGTTTGTATTTAATGAGTTAGTGACAACGTGGCCATTAAAAACCCCATCCGTTA  
ACCAAGCTCGTTTAGGTTTTGTTGCGCGTAATGCAAATGAAGCGATCGCGATGATTG  
ATACGGCATTGAAACAATTCAATGCGAACGCAGATAAAATGACATGGTCAGTACCTA  
CCGGGGTTTACTATCGTCAAGCCGGTATTGATGCAACAGGTAAAGTGGTTGCGCTAT  
TCTCAGGGCAAGGTTGCAATACGTGAACATGGGTCTGTAATTAACCTGTAACCTTCC  
CAAGCATGATGCACAGTGCTGCGGCGATGGATAAAGAGTTCAGTGCCGCTGGTTTAG  
GCCAGTTATCTGCAGTTACTTTCCCTATCCCTGTTTATACGGATGCCGAGCGTAAGC  
TACAAGAAGAGCAATTACGTTTAAACGCAACATGCGCAACCAGCGATTGGTAGTTTGA  
GTGTTGGTCTGTTCAAAACGTTTAAAGCAAGCAGGTTTTAAAGCTGATTTTGCTGCCG  
GTCATAGTTTCGGTGAGTTAACCGCATTATGGGCTGCCGATGTATTGAGCGAAAGCG  
ATTACATGATGTTAGCGCGTAGTCGTGGTCAAGCAATGGCTGCGCCAGAGCAACAAG  
ATTTTGATGCAGGTAAGATGGCCGCTGTTGTTGGTGATCCAAAGCAAGTCGCTGTGA  
TCATTGATACCCTTGATGATGTCTCTATTGCTAACTTCAACTCGAATAACCAAGTTG  
TTATTGCTGGTACTACGGAGCAGGTTGCTGTAGCGGTTACAACCTTAGGTAATGCTG  
GTTTCAAAGTTGTGCCACTGCCGGTATCTGCTGCGTTCCATACACCTTTAGTTTCGTC  
ACGCGCAAAAACCATTTGCTAAAGCGGTTGATAGCGCTAAATTTAAAGCGCCAAGCA  
TTCCAGTGTTTGCTAATGGCACAGGCTTGGTGCATTCAAGCAAACCGAATGACATTA

**FIG. 5-14**



72/134

AGAAAAACCTGAAAAACCACATGCTGGAATCTGTTCAATTCATCAAGAAATTGACA  
ACATCTATGCTGATGGTGGCCGCGTATTTATCGAATTTGGTCCAAAGAATGTATTAA  
CTAAATTGGTTGAAAACATTCTCACTGAAAAATCTGATGTGACTGCTATCGCGGTTA  
ATGCTAATCCTAAACAACCTGCGGACGTACAAATGCGCCAAGCTGCGCTGCAAATGG  
CAGTGCTTGGTGTGCGATTAGACAATATTGACCCGTACGACGCCGTTAAGCGTCCAC  
TTGTTGCGCCGAAAGCATCACCAATGTTGATGAAGTTATCTGCAGCGTCTTATGTTA  
GTCCGAAAACGAAGAAAGCGTTTGCTGATGCATTGACTGATGGCTGGACTGTTAAGC  
AAGCGAAAGCTGTACCTGCTGTTGTGTGTCACAACCACAAGTGATTGAAAAGATCGTTG  
AAGTTGAAAAGATAGTTGAACGCATTGTGCAAGTAGAGCGTATTGTGCAAGTAGAAA  
AAATCGTCTACGTTAATGCTGACGGTTCGCTTATATCGCAAATAATCAAGACGTTA  
ACAGCGCTGTTGTTAGCAACGTGACTAATAGCTCAGTGA CTCATAGCAGTGATGCTG  
ACCTTGTTGCCTCTATTGAACGCAGTGTTGGTCAATTTGTTGCACACCAACAGCAAT  
TATTAAATGTACATGAACAGTTTATGCAAGGTCCACAAGACTACGCGAAAACAGTG  
AGAACGTACTTGCTGCGCAGACGAGCAATGAATTACCGGAAAGTTTAGACCGTACAT  
TGTCTATGTATAACGAGTTCCAATCAGAAACGCTACGTGTACATGAAACGTACCTGA  
ACAATCAGACGAGCAACATGAACACCATGCTTACTGGTGCTGAAGCTGATGTGCTAG  
CAACCCCAATAACTCAGGTAGTGAATACAGCCGTTGCCACTAGTCACAAGGTAGTTG  
CTCCAGTTATTGCTAATACAGTGACGAATGTTGTATCTAGTGTGAGTAATAACGCGG  
CGGTTGCAGTGCAAACCTGTGGCATTAGCGCCTACGCAAGAAATCGCTCCAACAGTCG  
CTACTACGCCAGCACCCGCATTGGTTGCTATCGTGGCTGAACCTGTGATTGTTGCGC  
ATGTTGCTACAGAAGTTGCACCAATTACACCATCAGTTACACCAGTTGTGCGCAACTC  
AAGCGGCTATCGATGTAGCAACTATTAACAAAGTAATGTTAGAAGTTGTTGCTGATA  
AAACCGGTTATCCAACGGATATGCTGGAACCTGAGCATGGACATGGAAGCTGACTTAG  
GTATCGACTCAATCAAACGTGTTGAGATATTAGGCGCAGTACAGGAATTGATCCCTG  
ACTTACCTGAACTTAATCCTGAAGATCTTGCTGAGCTACGCACGCTTGGTGAGATTG  
TCGATTACATGAATTCAAAGCCCAGGCTGTAGCTCCTACAACAGTACCTGTAACAA

**FIG. 5-15**

73/134

GTGCACCTGTTTCGCCTGCATCTGCTGGTATTGATTTAGCCCAATCCAAAACGTAA  
TGTTAGAAGTGGTTGCAGACAAAACCGGTTACCCAACAGACATGCTAGAACTGAGCA  
TGGATATGGAAGCTGACTTAGGTATTGATTCAATCAAGCGTGTGGAAATCTTAGGTG  
CAGTACAGGAGATCATAACTGATTTACCTGAGCTAAACCCTGAAGATCTTGCTGAAT  
TACGCACCCTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCAGTCGCTGAAA  
GTGCGCCAGTGGCGACGGCTCCTGTAGCAACAAGCTCAGCACCGTCTATCGATTGTA  
ACCACATTCAAACAGTGATGATGGATGTAGTTGCAGATAAGACTGGTTATCCAACTG  
ACATGCTAGAACTTGGCATGGACATGGAAGCTGATTTAGGTATCGATTCAATCAAAC  
GTGTGGAAATATTAGGCGCAGTGCAGGAGATCATCACTGATTTACCTGAGCTAAACC  
CAGAAGACCTCGCTGAATTACGCACGCTAGGTGAAATCGTTAGTTACATGCAAAGCA  
AAGCGCCAGTCGCTGAGAGTGCGCCAGTAGCGACGGCTTCTGTAGCAACAAGCTCTG  
CACCGTCTATCGATTTAAACCATATCCAAACAGTGATGATGGAAGTGGTTGCAGACA  
AAACCGGTTATCCAGTAGACATGTTAGAACTTGCTATGGACATGGAAGCTGACCTAG  
GTATCGATTCAATCAAGCGTGTAGAAATTTTAGGTGCGGTACAGGAAATCATTACTG  
ACTTACCTGAGCTTAACCCTGAAGATCTTGCTGAACTACGTACATTAGGTGAAATCG  
TTAGTTACATGCAAAGCAAAGCGCCCGTAGCTGAAGCGCCTGCAGTACCTGTTGCAG  
TAGAAAGTGCACCTACTAGTGTAACAAGCTCAGCACCGTCTATCGATTTAGACCACA  
TCCAAAATGTAATGATGGATGTTGTTGCTGATAAGACTGGTTATCCTGCCAATATGC  
TTGAATTAGCAATGGACATGGAAGCCGACCTTGGTATTGATTCAATCAAGCGTGTTG  
AAATTCTAGGCGCGGTACAGGAGATCATTACTGATTTACCTGAACTAAACCCAGAAG  
ACTTAGCTGAACTACGTACGTTAGAAGAAATTGTAACCTACATGCAAAGCAAGGCGA  
GTGGTGTTACTGTAAATGTAGTGGCTAGCCCTGAAAATAATGCTGTATCAGATGCAT  
TTATGCAAAGCAATGTGGCGACTATCACAGCGGCCGCAGAACATAAGGCGGAATTTA  
AACCGGCGCCGAGCGCAACCGTTGCTATCTCTCGTCTAAGCTCTATCAGTAAAATAA  
GCCAAGATTGTAAAGGTGCTAACGCCTTAATCGTAGCTGATGGCACTGATAATGCTG

**FIG. 5-16**

74/134

TGTTACTTGCAGACCACCTATTGCAAACCTGGCTGGAATGTAACCTGCATTGCAACCAA  
CTTGGGTAGCTGTAACAACGACGAAAGCATTTAATAAGTCAGTGAACCTGGTGACTT  
TAAATGGCGTTGATGAACTGAAATCAACAACATTATTACTGCTAACGCACAATTGG  
ATGCAGTTATCTATCTGCACGCAAGTAGCGAAATTAATGCTATCGAATACCCACAAG  
CATCTAAGCAAGGCCTGATGTTAGCCTTCTTATTAGCGAAATTGAGTAAAGTAACTC  
AAGCCGCTAAAGTGCGTGGCGCCTTTATGATTGTTACTCAGCAGGGTGGTTCATTAG  
GTTTTGATGATATCGATTCTGCTACAAGTCATGATGTGAAAACAGACCTAGTACAAA  
GCGGCTTAAACGGTTTAGTTAAGACACTGTCTCACGAGTGGGATAACGTATTCTGTC  
GTGCGGTTGATATTGCTTCGTCATTAACGGCTGAACAAGTTGCAAGCCTTGTTAGTG  
ATGAACTACTTGATGCTAACACTGTATTAACAGAAGTGGGTATCAACAAGCTGGTA  
AAGGCCTTGAACGTATCACGTAACTGGTGTGGCTACTGACAGCTATGCATTAACAG  
CTGGCAATAACATCGATGCTAACTCGGTATTTTTAGTGAGTGGTGGCGCAAAGGTG  
TAACTGCACATTGTGTTGCTCGTATAGCTAAAGAATATCAGTCTAAGTTCATCTTAT  
TGGGACGTTCAACGTTCTCAAGTGACGAACCGAGCTGGGCAAGTGGTATTACTGATG  
AAGCGGCGTTAAAGAAAGCAGCGATGCAGTCTTTGATTACAGCAGGTGATAAACCAA  
CACCCGTAAAGATCGTACAGCTAATCAAACCAATCCAAGCTAATCGTGAAATTGCGC  
AAACCTTGTCTGCAATTACCGCTGCTGGTGGCCAAGCTGAATATGTTTCTGCAGATG  
TAACTAATGCAGCAAGCGTACAAATGGCAGTCGCTCCAGCTATCGCTAAGTTCGGTG  
CAATCACTGGCATCATTGATGGCGCGGGTGTGTTAGCTGACCAATTCATTGAGCAAA  
AAACACTGAGTGATTTTGAGTCTGTTTACAGCACTAAAATTGACGGTTTGTTATCGC  
TACTATCAGTCACTGAAGCAAGCAACATCAAGCAATTGGTATTGTTCTCGTCAGCGG  
CTGGTTTCTACGGTAACCCCGGCCAGTCTGATTACTCGATTGCCAATGAGATCTTAA  
ATAAAACCGCATACCGCTTTAAATCATTGCACCCACAAGCTCAAGTATTGAGCTTTA  
ACTGGGGTCCTTGGGACGGTGGCATGGTAACGCCTGAGCTTAAACGTATGTTTGACC  
AACGTGGTGTTTACATTATTCCACTTGATGCAGGTGCACAGTTATTGCTGAATGAAC

**FIG. 5-17**

75/134

TAGCCGCTAATGATAACCGTTGTCCACAAATCCTCGTGGGTAATGACTTATCTAAAG  
ATGCTAGCTCTGATCAAAAGTCTGATGAAAAGAGTACTGCTGTAAAAAGCCACAAG  
TTAGTCGTTTATCAGATGCTTTAGTAACTAAAAGTATCAAAGCGACTAACAGTAGCT  
CTTTATCAAACAAGACTAGTGCTTTATCAGACAGTAGTGCTTTTCAGGTAAACGAAA  
ACCACTTTTTAGCTGACCACATGATCAAAGGCAATCAGGTATTACCAACGGTATGCG  
CGATTGCTTGGATGAGTGATGCAGCAAAAGCGACTTATAGTAACCGAGACTGTGCAT  
TGAAGTATGTCGGTTTTCGAAGACTATAAATTGTTTAAAGGTGTGGTTTTTGTATGGCA  
ATGAGGCGGCGGATTACCAAATCCAATTGTGCGCTGTGACAAGGGCGTCAGAACAGG  
ATTCTGAAGTCCGTATTGCCGCAAAGATCTTTAGCCTGAAAAGTGACGGTAAACCTG  
TGTTTCATTATGCAGCGACAATATTGTTAGCAACTCAGCCACTTAATGCTGTGAAGG  
TAGAACTTCCGACATTGACAGAAAGTGTTGATAGCAACAATAAAGTAACTGATGAAG  
CACAAGCGTTATACAGCAATGGCACCTTGTTCCACGGTGAAAGTCTGCAGGGCATTA  
AGCAGATATTAAGTTGTGACGACAAGGGCCTGCTATTGGCTTGTGAGATAACCGATG  
TTGCAACAGCTAAGCAGGGATCCTTCCCGTTAGCTGACAACAATATCTTTGCCAATG  
ATTTGGTTTTATCAGGCTATGTTGGTCTGGGTGCGCAAACAATTTGGTTTAGGTAGCT  
TACCTTCGGTGACAACGGCTTGGACTGTGTATCGTGAAGTGGTTGTAGATGAAGTAT  
TTTATCTGCAACTTAATGTTGTTGAGCATGATCTATTGGGTTACGCGGCAGTAAAG  
CCCGTTGTGATATTCAATTGATTGCTGCTGATATGCAATTACTTGCCGAAGTGAAAT  
CAGCGCAAGTCAGTGTGAGTGACATTTTGAACGATATGTCATGATCGAGTAAATAAT  
AACGATAGGCGTCATGGTGAGCATGGCGTCTGCTTCTTCATTTTTTAACATTAAACA  
ATATTAATAGCTAAACGCGGTTGCTTTAAACCAAGTAAACAAGTGCTTTTAGCTATT  
ACTATTCCAAACAGGATATTAAAGAGAATATGACGGAATTAGCTGTTATTGGTATGG  
ATGCTAAATTTAGCGGACAAGACAATATTGACCGTGTGGAACGCGCTTCTATGAAG  
GTGCTTATGTAGGTAATGTTAGCCGCGTTAGTACCGAATCTAATGTTATTAGCAATG  
GCGAAGAACAAGTTATTACTGCCATGACAGTTCCTTAACTCTGTGAGTCTACTAGCGC

**FIG. 5-18**

76/134

AAACGAATCAGTTAAATATAGCTGATATCGCGGTGTTGCTGATTGCTGATGTAAAAA  
GTGCTGATGATCAGCTTGTAGTCCAAATTGCATCAGCAATTGAAAAACAGTGTGCGA  
GTTGTGTTGTTATTGCTGATTTAGGCCAAGCATTAAATCAAGTAGCTGATTTAGTTA  
ATAACCAAGACTGTCTGTGGCTGTAATTGGCATGAATAACTCGGTTAATTTATCTC  
GTCATGATCTTGAATCTGTAAGTCAACAATCAGCTTTGATGAAACCTTCAATGGTT  
ATAACAATGTAGCTGGGTTCGCGAGTTTACTTATCGCTTCAACTGCGTTTGCCAATG  
CTAAGCAATGTTATATATACGCCAACATTAAGGGCTTCGCTCAATCGGGCGTAAATG  
CTCAATTTAACGTTGGAAACATTAGCGATACTGCAAAGACCGCATTGCAGCAAGCTA  
GCATAACTGCAGAGCAGGTTGGTTTGTAGAAAGTGTGAGCAGTCGCTGATTCGGCAA  
TCGCATTGTCTGAAAGCCAAGGTTTAAATGTCTGCTTATCATCATACGCAAACCTTTGC  
ATACTGCATTAAGCAGTGCCCGTAGTGTGACTGGTGAAGGCGGGTGTTTTTACAGG  
TCGCAGGTTTATTGAAATGTGTAATTGGTTTACATCAACGTTATATTCCGGCGATTA  
AAGATTGGCAACAACCGAGTGACAATCAAATGTCACGGTGGCGGAATTCACCATTCT  
ATATGCCTGTAGATGCTCGACCTTGGTTCCACATGCTGATGGCTCTGCACACATTG  
CCGCTTATAGTTGTGTGACTGCTGACAGCTATTGTCATATTCTTTTACAAGAAAACG  
TCTTACAAGAACTTGTTTTGAAAGAAACAGTCTTGCAAGATAATGACTTAACTGAAA  
GCAAGCTTCAGACTCTTGAACAAAACAATCCAGTAGCTGATCTGCGCACTAATGGTT  
ACTTTGCATCGAGCGAGTTAGCATTAAATCATAGTACAAGGTAATGACGAAGCACAAT  
TACGCTGTGAATTAGAACTATTACAGGGCAGTTAAGTACTACTGGCATAAGTACTA  
TCAGTATTAAACAGATCGCAGCAGACTGTTATGCCCGTAATGATACTAACAAAGCCT  
ATAGCGCAGTGCTTATTGCCGAGACTGCTGAAGAGTTAAGCAAAGAAATAACCTTGG  
CGTTTGCTGGTATCGCTAGCGTGTTTAAATGAAGATGCTAAAGAATGGAAAACCCCGA  
AGGGCAGTTATTTTACCGCGCAGCCTGCAAATAAACAGGCTGCTAACAGCACACAGA  
ATGGTGTCACCTTCATGTACCCAGGTATTGGTGCTACATATGTTGGTTTAGGGCGTG  
ATCTATTTTCATCTATTCCACAGATTTATCAGCCTGTAGCGGCTTTAGCCGATGACA

**FIG. 5-19**

77/134

TTGGCGAAAGTCTAAAAGATACTTTACTTAATCCACGCAGTATTAGTCGTCATAGCT  
TTAAAGAACTCAAGCAGTTGGATCTGGACCTGCGCGGTAACCTTAGCCAATATCGCTG  
AAGCCGGTGTGGGTTTTGCTTGTGTGTTTACCAAGGTATTTGAAGAAGTCTTTGCCG  
TTAAAGCTGACTTTGCTACAGGTTATAGCATGGGTGAAGTAAGCATGTATGCAGCAC  
TAGGCTGCTGGCAGCAACCGGGATTGATGAGTGCTCGCCTTGCAACAATCGAATACCT  
TTAATCATCAACTTTGCGGCGAGTTAAGAACACTACGTCAGCATTGGGGCATGGATG  
ATGTAGCTAACGGTACGTTTCGAGCAGATCTGGGAAACCTATAACCATTAAGGCAACGA  
TTGAACAGGTCGAAATTGCCTCTGCAGATGAAGATCGTGTGTATTGCACCATTATCA  
ATACACCTGATAGCTTGTTGTTAGCCGGTTATCCAGAAGCCTGTCAGCGAGTCATTA  
AGAATTTAGGTGTGCGTGCAATGGCATTGAATATGGCGAACGCAATTCACAGCGCGC  
CAGCTTATGCCGAATACGATCATATGGTTGAGCTATACCATATGGATGTTACTCCAC  
GTATTAATACCAAGATGTATTCAAGCTCATGTTATTTACCGATTCCACAACGCAGCA  
AAGCGATTTCCACAGTATTGCTAAATGTTTGTGTGATGTGGTGGATTTCCACGTT  
TGGTTAATACCTTACATGACAAAGGTGCGCGGGTATTCATTGAAATGGGTCCAGGTC  
GTTTCGTTATGTAGCTGGGTAGATAAGATCTTAGTTAATGGCGATGGCGATAATAAAA  
AGCAAAGCCAACATGTATCTGTTCCGTGTGAATGCCAAAGGCACCAGTGATGAACTTA  
CTTATATTCGTGCGATTGCTAAGTTAATTAGTCATGGCGTGAATTTGAATTTAGATA  
GCTTGTTTAAACGGGTCAATCCTGGTTAAAGCAGGCCATATAGCAAACACGAACAAAT  
AGTCAACATCGATATCTAGCGCTGGTGAGTTATACCTCATTAGTTGAAATATGGATT  
TAAAGAGAGTAATTATGGAAAATATTGCAGTAGTAGGTATTGCTAATTTGTTCCCGG  
GCTCACAAGCACCGGATCAATTTTGGCAGCAATTGCTTGAACAACAAGATTGCCGCA  
GTAAGGCGACCGCTGTTCAAATGGGCGTTGATCCTGCTAAATATACCGCCAACAAAG  
GTGACACAGATAAATTTTACTGTGTGCACGGCGGTTACATCAGTGATTTCAATTTTG  
ATGCTTCAGGTTATCAACTCGATAATGATTATTTAGCCGGTTTAGATGACCTTAATC  
AATGGGGGCTTTATGTTACGAAACAAGCCCTTACCGATGCGGGTATTGGGGCAGTA

**FIG. 5-20**

78/134

CTGCACTAGAAAACGTGGTGTGATTTTAGGTAATTTGTCATTCCCAACTAAATCAT  
CTAATCAGCTGTTTATGCCTTTGTATCATCAAGTTGTTGATAATGCCTTAAAGGCGG  
TATTACATCCTGATTTTCAATTAACGCATTACACAGCACCGAAAAAACACATGCTG  
ACAATGCATTAGTAGCAGGTTATCCAGCTGCATTGATCGCGCAAGCGGCGGGTCTTG  
GTGGTTCACATTTTGCACCTGGATGCGGCTTGTGCTTCATCTTGTTATAGCGTTAAGT  
TAGCGTGTGATTACCTGCATACGGGTAAAGCCAACATGATGCTTGCTGGTGCGGTAT  
CTGCAGCAGATCCTATGTTTCGTAAATATGGGTTTCTCGATATTCGAAGCTTACCCAG  
CTAACAATGTACATGCCCCGTTTGACCAAAATTACAAGGTCTATTTGCCGGTGAAG  
GCGCGGGCATGATGGTATTGAAACGTCAAAGTGATGCAGTACGTGATGGTGATCATA  
TTTACGCCATTATTAAAGGCGGCGCATTATCGAATGACGGTAAAGGCGAGTTTGTAT  
TAAGCCCGAACACCAAGGGCCAAGTATTAGTATATGAACGTGCTTATGCCGATGCAG  
ATGTTGACCCGAGTACAGTTGACTATATTGAATGTCATGCAACGGGCACACCTAAGG  
GTGACAATGTTGAATTGCGTTCGATGGAAACCTTTTTCAGTCGCGTAAATAACAAAC  
CATTACTGGGCTCGGTAAATCTAACCTTGGTCATTTGTTAACTGCCGCTGGTATGC  
CTGGCATGACCAAAGCTATGTTAGCGCTAGGTAAAGGTCTTATTCCTGCAACGATTA  
ACTTAAAGCAACCACTGCAATCTAAAAACGGTTACTTTACTGGCGAGCAAATGCCAA  
CGACGACTGTGTCTTGGCCAACAACCTCCGGGTGCCAAGGCAGATAAACCGCGTACCG  
CAGGTGTGAGCGTATTTGGTTTTGGTGGCAGCAACGCCCATTGTTATTACAACAGC  
CAACGCAAACACTCGAGACTAATTTTAGTGTGCTAAACCACGTGAGCCTTTGGCTA  
TTATTGGTATGGACAGCCATTTTGGTAGTGCCAGTAATTTAGCGCAGTTCAAAACCT  
TATTAAATAATAATCAAAATACCTCCGTGAATTACCAGAACAACGCTGGAAAGGCA  
TGGAAAGTAACGCTAACGTCATGCAGTCGTTACAATTACGCAAAGCGCCTAAAGGCA  
GTTACGTTGAACAGCTAGATATTGATTTCTTGCGTTTTAAAGTACCGCCTAATGAAA  
AAGATTGCTTGATCCCGCAACAGTTAATGATGATGCAAGTGGCAGACAATGCTGCGA  
AAGACGGAGGTCTAGTTGAAGGTCGTAATGTTGCGGTATTAGTAGCGATGGGCATGG

**FIG. 5-21**

79/134

AACTGGAATTACATCAGTATCGTGGTCGCGTTAATCTAACCACCCAAATTGAAGACA  
GCTTATTACAGCAAGGTATTAACCTGACTGTTGAGCAACGTGAAGAACTGACCAATA  
TTGCTAAAGACGGTGTTCCTCGGCTGCACAGCTAAATCAGTATACGAGTTTCATTG  
GTAATATTATGGCGTCACGTATTTTCGGCGTTATGGGATTTTTCTGGTCCTGCTATTA  
CCGTATCGGCTGAAGAAACTCTGTTTATCGTTGTGTTGAATTAGCTGAAAATCTAT  
TTCAAACCACTGATGTTGAAGCCGTTATTATTGCTGCTGTTGATTTGTCTGTTCAA  
TTGAAACATTACTTTACGTCAGCACTACGGTCCAGTTAATGAAAAGGGATCTGTAA  
GTGAATGTGGTCCGGTTAATGAAAGCAGTTCAGTAACCAACAATATTCTTGATCAGC  
AACAAATGGCTGGTGGGTGAAGGCGCAGCGGCTATTGTCGTTAAACCGTCATCGCAAG  
TCACTGCTGAGCAAGTTTATGCGCGTATTGATGCGGTGAGTTTTGCCCTGGTAGCA  
ATGCGAAAGCAATTACGATTGCAGCGGATAAAGCATTAACTTGCTGGTATCAGTG  
CTGCTGATGTAGCTAGTGTGAAGCACATGCAAGTGGTTTTAGTGCCGAAAATAATG  
CTGAAAAAACCGGTTACCGACTTTATACCCAAGCGCAAGTATCAGTTCGGTGAAAG  
CCAATATTGGTCATACGTTTAAATGCCTCGGGTATGGCGAGTATTATTAAACGGCGC  
TGCTGTTAGATCAGAATACGAGTCAAGATCAGAAAAGCAAACATATTGCTATTAACG  
GTCTAGGTCGTGATAACAGCTGCGCGCATCTTATCTTATCGAGTTCAGCGCAAGCGC  
ATCAAGTTGCACCAGCGCCTGTATCTGGTATGGCCAAGCAACGCCCACAGTTAGTTA  
AAACCATCAAACCTCGGTGGTCAGTTAATTAGCAACGCGATTGTTAACAGTGCAGATT  
CATCTTTACACGCTATTAAAGCGCAGTTTGCCGGTAAGCACTTAAACAAAGTTAACC  
AGCCAGTGATGATGGATAACCTGAAGCCCCAAGGTATTAGCGCTCATGCAACCAATG  
AGTATGTGGTGACTGGAGCTGCTAACACTCAAGCTTCTAACATTCAAGCATCTCATG  
TTCAAAGCGTCAAGTCATGCACAAGAGATAGCACCAAAACCAAGTTCAAATATGCAAG  
CTACAGCAGCCGCTGTAAGTTCACCCCTTTCTCAACATCAACACACAGCGCAGCCCG  
TAGCGGCACCGAGCGTTGTTGGAGTGAAGTGTGAAACATAAAGCAAGTAACCAAATTC  
ATCAGCAAGCGTCTACGCATAAAGCATTTTTAGAAAGTCGTTTAGCTGCACAGAAAA

**FIG. 5-22**



80/134

ACCTATCGCAACTTGTTGAATTGCAAACCAAGCTGTCAATCCAAACTGGTAGTGACA  
ATACATCTAACAATACTGCGTCAACAAGCAATACAGTGCTAACAAATCCTGTATCAG  
CAACGCCATTAACTTGTGTCTAATGCGCCTGTAGTAGCGACAAACCTAACCAGTA  
CAGAAGCAAAAGCGCAAGCAGCTGCTACACAAGCTGGTTTTTCAGATAAAAGGACCTG  
TTGGTTACAACCTATCCACCGCTGCAGTTAATTGAACGTTATAATAAACAGAAAACG  
TGATTTACGATCAAGCTGATTTGGTTGAATTGCTGAAGGTGATATTGGTAAGGTAT  
TTGGTGCTGAATACAATATTATTGATGGCTATTCGCGTCGTGTACGTCTGCCAACCT  
CAGATTACTTGTTAGTAACACGTGTTACTGAACTTGATGCCAAGGTGCATGAATACA  
AGAAATCATACTGTGTACTGAATATGATGTGCCTGTTGATGCACCGTTCTTAATTG  
ATGGTCAGATCCCTTGGTCTGTTGCCGTCGAATCAGGCCAGTGTGATTTGATGTTGA  
TTTCATATATCGGTATTGATTTCCAAGCGAAAGGCGAACGTGTTTACCGTTTACTTG  
ATTGTGAATTAACTTTCCTTGAAGAGATGGCTTTTGGTGCGGATACTTTACGTTACG  
AGATCCACATTGATTTCGTATGCACGTAACGGCGAGCAATTATTATTCTTCTTCCATT  
ACGATTGTTACGTAGGGGATAAGAAGGTACTTATCATGCGTAATGGTTGTGCTGGTT  
TCTTTACTGACGAAGAACTTTCTGATGGTAAAGGCGTTATTCATAACGACAAAGACA  
AAGCTGAGTTTAGCAATGCTGTAAATCATCATTACGCCGTTATTACAACATAACC  
GTGGTCAATACGATTATAACGACATGATGAAGTTGGTTAATGGTGATGTTGCCAGTT  
GTTTTGGTCCGCAATATGATCAAGGTGGCCGTAATCCATCATTGAAATTCTCGTCTG  
AGAAGTTCTTGATGATTGAACGTATTACCAAGATAGACCCAACCGGTGGTCATTGGG  
GACTAGGCCTGTTAGAAGGTCAGAAAGATTTAGACCCTGAGCATTGGTATTTCCCTT  
GTCACTTTAAAGGTGATCAAGTAATGGCTGGTTCGTTGATGTCGGAAGGTTGTGGCC  
AAATGGCGATGTTCTTCATGCTGTCTCTTGGTATGCATACCAATGTGAACAACGCTC  
GTTTCCAACCACTACCAGGTGAATCACAACGGTACGTTGTCGTGGGCAAGTACTGC  
CACAGCGCAATACCTTAACTTACCGTATGGAAGTTACTGCGATGGGTATGCATCCAC  
AGCCATTTCATGAAAGCTAATATTGATATTTTGCTTGACGGTAAAGTGGTTGTTGATT

FIG. 5-23

81/134

TCAAAACTTGAGCGTGATGATCAGCGAACAAGATGAGCATTACAGATTACCTGTAA  
CACTGCCGAGTAATGTGGCGCTTAAAGCGATTACTGCACCTGTTGCGTCAGTAGCAC  
CAGCATCTTCACCCGCTAACAGCGCGGATCTAGACGAACGTGGTGTGTAACCGTTTA  
AGTTTCCTGAACGTCCGTTAATGCGTGTTGAGTCAGACTTGTCTGCACCGAAAAGCA  
AAGGTGTGACACCGATTAAAGCATTTTGAAGCGCCTGCTGTTGCTGGTCATCATAGAG  
TGCCTAACCAAGCACCGTTTACACCTTGGCATATGTTTGAGTTTTCGACGGGTAATA  
TTTCTAACTGTTTCGGTCCTGATTTTGATGTTTATGAAGGTCGTATTCCACCTCGTA  
CACCTTGTGGCGATTTACAAGTTGTTACTCAGGTTGTAGAAGTGCAGGGCGAACGTC  
TTGATCTTAAAAATCCATCAAGCTGTGTAGCTGAATACTATGTACCGGAAGACGCTT  
GGTACTTTACTAAAAACAGCCATGAAAACCTGGATGCCTTATTCATTAATCATGGAAA  
TTGCATTGCAACCAAATGGCTTTATTTCTGGTTACATGGGCACGACGCTTAAATACC  
CTGAAAAAGATCTGTTCTTCCGTAACCTTGATGGTAGCGGCACGTTATTAAAGCAGA  
TTGATTTACGCGGCAAGACCATTGTGAATAAATCAGTCTTGGTTAGTACGGCTATTG  
CTGGTGGCGCGATTATTCAAAGTTTCACGTTTGATATGTCTGTAGATGGCGAGCTAT  
TTTATACTGGTAAAGCTGTATTTGGTTACTTTAGTGGTGAATCACTGACTAACCAAC  
TGGGCATTGATAACGGTAAAACGACTAATGCGTGTTTTGTTGATAACAATACCCCG  
CAGCGAATATTGATGTGTTTGATTTAACTAATCAGTCATTGGCTCTGTATAAAGCGC  
CTGTGGATAAACCGCATTATAAATTGGCTGGTGGTCAGATGAACTTTATCGATACAG  
TGTCAGTGGTTGAAGGCGGTGGTAAAGCGGGCGTGGCTTATGTTTATGGCGAACGTA  
CGATTGATGCTGATGATTGGTTCTTCCGTTATCACTTCCACCAAGATCCGGTGATGC  
CAGGTTCAATTAGGTGTTGAAGCTATTATTGAGTTGATGCAGACCTATGCGCTTAAAA  
ATGATTTGGGTGGCAAGTTTGCTAACCCACGTTTTATTGCGCCGATGACGCAAGTTG  
ATTGGAAATACCGTGGGCAAATTACGCCGCTGAATAAACAGATGTCACTGGACGTGC  
ATATCACTGAGATCGTGAATGACGCTGGTGAAGTGCGAATCGTTGGTGATGCGAATC  
TGTCTAAAGATGGTCTGCGTATTTATGAAGTTAAAAACATCGTTTTAAGTATTGTTG

**FIG. 5-24**

82/134

AAGCGTAAAGGGTCAAGTGTAACGTGCTTAAGCGCCGCATTGGTTAAAGACGCTTTG  
CACGCCGTGAATCCGTCCATGGAGGCTTGGGGTTGGCATCCATGCCAACAACAGCAA  
GCTTACTTTAATCAATACGGCTTGGTGTCCATTTAGACGCCTCGAACTTAGTAGTTA  
ATAGACAAAATAATTTAGCTGTGGAATGAATATAGTAAGTAATCATTCGGCAGCTAC  
AAAAAAGGAATTAAGAATGTCGAGTTTAGGTTTTAACAATAACAACGCAATTAAGT  
GGCTTGGAAGTAGATCCAGCGTCAGTTCATACACAAGATGCAGAAATTAAAGCAGC  
TTTAATGGATCTAACTAAACCTCTCTATGTGGCGAATAATTCAGGCGTAACTGGTAT  
AGCTAATCATACGTCAGTAGCAGGTGCGATCAGCAATAACATCGATGTTGATGTATT  
GGCGTTTGCGCAAAAGTTAAACCCAGAAGATCTGGGTGATGATGCTTACAAGAAACA  
GCACGGCGTTAAATATGCTTATCATGGCGGTGCGATGGCAAATGGTATTGCCTCGGT  
TGAATTGGTTGTTGCGTTAGGTAAAGCAGGGCTGTTATGTTTCATTTGGTGCTGCAGG  
TCTAGTGCCTGATGCGGTTGAAGATGCAATTCGTCTGATTCAAGCTGAATTACCAAA  
TGCCCTTATGCGGTTAACTTGATCCATGCACCAGCAGAAGAAGCATTAGAGCGTG  
CGCGGTTGAACGTTTCCTAAAACTTGCGGTCAAGACGGTAGAGGCTTCAGCTTACCT  
TGGTTTAACTGAACACATTGTTTGGTATCGTGCTGCTGGTCTAACTAAAAACGCAGA  
TGGCAGTGTTAATATCGGTAACAAGGTTATCGCTAAAGTATCGCGTACCGAAGTTGG  
TCGCCGCTTTATGGAACCTGCACCGCAAAAATTACTGGATAAGTTATTAGAACAAAA  
TAAGATCACCCCTGAACAAGCTGCTTTAGCGTTGCTTGTACCTATGGCTGATGATAT  
TACTGGGGAAGCGGATTCTGGTGGTCATACAGATAACCGTCCGTTTTTAACATTATT  
ACCGACGATTATTGGTCTGCGTGATGAAGTGCAAGCGAAGTATAACTTCTCTCCTGC  
ATTACGTGTTGGTGCTGGTGGTGGTATCGGAACGCCTGAAGCAGCACTCGCTGCATT  
TAACATGGGCGCGGCTTATATCGTTCTGGGTTCTGTGAATCAGGCGTGTGTTGAAGC  
GGGTGCATCTGAATATACTCGTAACTGTTATCGACAGTTGAAATGGCTGATGTGAC  
TATGGCACCTGCTGCAGATATGTTTGAAATGGGTGTGAAGCTGCAAGTATTAACG  
CGGTTCTATGTTGCGATGCGTGCGAAGAACTGTATGACTTGTATGTGGCTTATGA

**FIG. 5-25**

83/134

CTCGATTGAAGATATCCCAGCTGCTGAACGTGAGAAGATTGAAAAACAAATCTTCCG  
TGCAAACCTAGACGAGATTTGGGATGGCACTATCGCTTTCTTTACTGAACGCGATCC  
AGAAATGCTAGCCCGTGCAACGAGTAGTCCTAAACGTAAAATGGCACTTATCTTCCG  
TTGGTATCTTGGCCTTTCTTCACGCTGGTCAAACACAGGCGAGAAGGGACGTGAAAT  
GGATTATCAGATTTGGGCAGGCCCAAGTTTAGGTGCATTCAACAGCTGGGTGAAAGG  
TTCTTACCTTGAAGACTATACCCGCCGTGGCGCTGTAGATGTTGCTTTGCATATGCT  
TAAAGGTGCTGCGTATTTACAACGTGTAAACCAGTTGAAATTGCAAGGTGTTAGCTT  
AAGTACAGAATTGGCAAGTTATCGTACGAGTGATTAATGTTACTTGATGATATGTGA  
ATTAATTAAAGCGCCTGAGGGCGCTTTTTTTGGTTTTTA ACTCAGGTGTTGTA ACTC  
GAAATTGCCCTTTCAAGTTAGATCGATTACTCACTCACAATATGTTGATATCGCAC  
TTGCCATATACTTGCTCATCCAAAGCCCTATATTGATAATGGTGTTAATAGTCTTTA  
ATATCCGAGTCTTTCTTCAGCATAATACTAATATAGAGACTCGACCAATGTTAAACA  
CÀACAAAGAATATATTCTTGTGTACTGCCTTATTATTAACGAGTGCGAGTACGACAG  
CTACTACGCTAAACAATTCGATATCAGCAATTGAACAACGTATTTCTGGTCGTATCG  
GTGTGGCTGTTTTAGATACGCAAATAAACAAACGTGGGCTTACAATGGTGATGCAC  
ATTTTCCGATGATGAGTACATTCAAAACCCTCGCTTGCGCGAAAATGCTAAGTGAAT  
CGACAAATGGTAATCTGGATCCAGTACTAGCTCATTGATAAAGGCTGAAGAATTAA  
TCCCTTGGTCACCAGTCACTAAAACGTTTGTGAATAACACTATTACAGTGGCGAAAG  
CGTGTGAAGCAACAATGCTGACCAGTGATAATACCGCGGCTAATATTGTTTTACAGT  
ATATCGGAGGCCCTCAAGGCGTTACTGCATTCTTGCGAGAAATTGGTGATGAAGAGA  
GTCAGTTAGATCGTATAGAACCTGAATTGAATGAAGCTAAGGTGCGAGACTTGCGTG  
ATACCACGACACCGAAAGCCATAGTTACCACGCTCAACAACTACTACTTGGTGATG  
TTCTACTTGATTTGGATAAAAACCAACTTAAAACATGGATGCAAATAATAAAGTGT  
CAGATCCTTTACTGCGTCTATATTACCGCAAGGCTGGTTTATTGCCGACCGCTCAG  
GTGCGGGTGGTAATGGTTCTCGAGGTATAACTGCTATGCTTTGGCACTCCGAGCGTC

**FIG. 5-26**

84/134

AACCGCTAATCATCAGTATTTATTTAACCGAACTGAGTTAGCAATGGCAATGCGCA  
ATGAGATTATTGTTGAGATCGGTAAGCTGATATTCAAAGAATACGCGGTGAAATAAT  
AAGTTATTTTTTGATAATACTTTAACGAGCGTAGCTATCGAAGTGAGGGCGTCAATT  
AGACACCTTTGCTTCCCCTACAAAATCTAATGTGTATTACCTCGGCTAGTACAATTG  
CCCTAAGTTATTTCTGTCCAGCTTTGGCTTAGTGCAATTGCGTTAGCCAATGTGAAC  
ACCAAGGGACTTTGTCGTACCATAACTACCAAGCGACTTTGTCGTTTTATCTTTTC  
TTAGACAAACAGAGGTTAAATGAGTGACGCCTTCCAAATCACAGGAATGAATCCGCA  
TTTCAATAAAATCTAACCCGTACCAACTCCGTACAAGTTGATCTTTAGTTGTTTAA  
ATCTATAATAAATTCAATTACGGAATTAATCCGTACAACCTGGAGGTTTTATGGCTAC  
TGCAAGACTTGATATCCGTTTGGATGAAGAAATCAAAGCTAAGGCTGAGAAAGCATC  
AGCTTTACTCGGCTTAAAAAGTTTAAACCGAATACGTTGTTTCGCTTAATGGACGAAGA  
TTCAACTAAAGTAGTTTCTGAGCATGAGAGTATTACCGTTGAAGCGAATGTATTCGA  
CCAATTTATGGCTGCTTGTGATGAAGCGAAAGCCCCAAATAAAGCATTACTGAAGC  
CGCTGTATTTACTCAGAATGGTGAGTTTAAGTGAGTTATTCCAAACGTTTCAAAGAA  
CTGGATAAATCAAAACATGACAGAGCATCATTTGACTGTGGCGAAAAAGAGCTAAAT  
GATTTTATCCAACTCAAGCAGCCAAACATATGCAAGCAGGTATTAGCCGCACTCTG  
GTTTTACCTGCTTCTGCGCCGTTACCAACAAAAAATATCCAATTTGCTCATTTTAT  
AGTATCGCGCCAAGCTCAATTAGCCGCGATACGTTACCACAAGCAATGGCTAAAAAG  
TTACCACGTTATCCTATCCCTGTTTTCTTTTGGCTCAACTTGCCGTCCATAAAGAG  
TTTCATGGGAGTGGGTTAGGCAAAGTTAGCTTAATTAAAGCGTTAGAGTACCTTTGG  
GAAATTAACCTCTCACATGAGAGCTTACGCCATCGTTGTTGATTGTTTAACTGAACAA  
GCTGAGTCATTCTACGCTAAATATGGTTTCGACGTTCTCTGCGAAATAAATGGTCGA  
GTAAGAATGTTTCATATCAATGAAAACAGTCAATCAGTTATTCACTTAACAGTAAGAG  
TTAGTATAACAGTTGTATGAATTAAATTTATTATATTCGGTAATCTCATTGCGATCA  
CGCTAGAAGTGCGAGCGGGTCAGACCGAGGCCACAATAGCAGCCGTTACGTTTAGGG

FIG. 5-27

85/134

GATGACTTAAAAAGATAACTACTACGTCAGTGGCGATCCTAGAGGATTAAAGGTTTA  
TGATTACACAACATTTATTTATTGTGCTTAATTTTTTCTATCCAATATGCGCAAGCTG  
TAAATATCACTGAAGTAGACTTTTATGTCAGTGATGATATCCCTAAAGATGTTGCCA  
AATTAAAGATAGGTGAATCCATAACGAACTCCAGCCTTATTCTAAGTAACTCATCTA  
TTCCACTCTCGCGGGAGACGGGTAACATATATTACTCTTCATCAATTGCTAACTTGA  
ACTATGACTCGATAGAATTTGTTATGGCTCAATTGATGGCCGAAGATTCCAGCCTTT  
ACAAGATGCTGGTAAATAGCGATAGGTTGTCCGTGCTAGTAATGACATCTTCCCAGT  
CCACAGATCTCTATGGCTCGACTTACTCGGCTTATTTTCCTAATGTTGCGGTCATCG  
ATTTGAATTGTGACTCGCTAACTTTAGAACATGAGCTCGGCCATCTATACGGAGCTG  
AACATGAAGAAATATATGACGACTATGTCTTCTATGCTGCGATATGTGGAGACTATA  
CGACTATCATGAACTCTATGCAGCCTGAAATGAAAGAAAAACAAATGATAAAGGCAT  
ATTCATTCCCTGAATTAAAAGTGGATGGCTTGCAAGTGGGAAATGAAAATACGAATA  
ACAAAAGGTTATTTTAGACAATATTGGTCGGTTTAGATAGGATTGGGATATTATTC  
TCATTGCGCTCTACTTAGTGCTGTTATTATGAGTGCCAGTGCTTCTATCTACGATAT  
TGGTCTTAACAAGTATTTATCTATAGACGCTAAGGTGTTATGTATTTAAGGGATGTT  
CAAGATGAACTAGGTGTAAACGATGTATAGTTGTATAACATTTTTTCAACGGTTGG  
AACGTTGATTCTATCGGGTAACAAGACCGCGACGATCCGCGATAAGTCCGATAGTC  
ATTACTTAGTTGGTCAGATGTTAGATGCTTGTAAGTACGGAAGATAATCGGAAAATGT  
GTCAAATAGAAATACTGAGCATTGAATATGTGACGTTTAGTGAATTAAACCGTGCGC  
ACGCCAATGCTGAAGGTTTACCGTTTTTGTATTATGCTTAAGTGGATAGTTGAAAGA  
TTTATCCGACTTCAAATGATTTATTTTTCATAAGTTTCAGAGTTGTAAGTATCGATA  
TCTTATAAGTCTTAGTGACAAAACAGAACTATTTATAGCGCTCAAGAAGGCGATAA  
TTTGATAATGAATTATCGCCTTGTTACTATTAAGAGACTTTAAATGACTGAGATATA  
AGATATGACACGGAAGAACATATTGATCACAGGCGCAAGTTCAGGGTTGGGCCGAGG  
TATGGCCATCGAATTTGCAAAATCAGGTCATAACTTAGCACTTTGTGCACGTAGACT

**FIG. 5-28**

86/134

TGATAATTTAGTTGCACTGAAAGCAGAACTCTTAGCCCTCAATCCTCACATCCAAAT  
CGAAATAAAACCTCTTGATGTCAATGAACATGAACAAGTCTTCACTGTTTTCCATGA  
ATTCAAAGCTGAATTGGTACGCTTGATCGTATTATTGTTAATGCTGGATTAGGCAA  
GGGTGGATCC

\*  
40138

**FIG. 5-29**

87/134

1

\*  
AAATGCAATTAATTATGGCGTAAATAGAGTGAAAACATGGCTAATATTCACCTAAGTC  
CTGAATTTTATATAAAGTTTAATCTGTTATTTTAGCGTTTACCTGGTCTTATCAGTG  
AGGTTTATAGCCATTATTAGTGGGATTGAAGTGATTTTTAAAGCTATGTATATTATT  
GCAAATATAAATTGTAACAATTAAGACTTTGGACACTTGAGTTCAATTTCGAATTGA  
TTGGCATAAAATTTAAACAGCTAAATCTACCTCAATCATTTTAGCAAATGTATGCA  
GGTAGATTTTTTTTCGCCATTTAAGAGTACACTTGACGCTAGGTTTTTGTGTTAGTGT  
GCAAATGAACGTTTTGATGAGCATTGTTTTTAGAGCACAAAATAGATCCTTACAGGA  
GCAATAACGCAATGGCTAAAAAGAACACCACATCGATTAAGCACGCCAAGGATGTGT  
TAAGTAGTGATGATCAACAGTTAAATTCTCGCTTGCAAGAATGTCCGATTGCCATCA  
TTGGTATGGCATCGGTTTTTGCAGATGCTAAAACTTGGATCAATTCTGGGATAACA  
TCGTTGACTCTGTGGACGCTATTATTGATGTGCCTAGCGATCGCTGGAACATTGACG  
ACCATTACTCGGCTGATAAAAAAGCAGCTGACAAGACATACTGCAAACGCGGTGGTT  
TCATTCCAGAGCTTGATTTTGATCCGATGGAGTTTGGTTTACCGCCAAATATCCTCG  
AGTTAACTGACATCGCTCAATTGTTGTCATTAATTGTTGCTCGTGATGTATTAAGTG  
ATGCTGGCATTGGTAGTGATTATGACCATGATAAAATTGGTATCACGCTGGGTGTGCG  
GTGGTGGTCAGAAACAAATTTTCGCCATTAACGTCGCGCCTACAAGGCCCGGTATTAG  
AAAAAGTATTAAAAGCCTCAGGCATTGATGAAGATGATCGCGCTATGATCATCGACA  
AATTTAAAAAGCCTACATCGGCTGGGAAGAGAACTCATTCCCAGGCATGCTAGGTA  
ACGTTATTGCTGGTCGTATCGCCAATCGTTTTGATTTTGGTGGTACTAACTGTGTGG  
TTGATGCGGCATGCGCTGGCTCCCTTGACAGCTGTTAAAATGGCGATCTCAGACTTAC  
TTGAATATCGTTCAGAAGTCATGATATCGGGTGGTGTATGTTGTGATAACTCGCCAT  
TCATGTATATGTCATTCTCGAAAACACCAGCATTTACCACCAATGATGATATCCGTC  
CGTTTGATGACGATTCAAAGGCATGCTGGTTGGTGAAGGTATTGGCATGATGGCGT  
TTAAACGTCCTGAAGATGCTGAACGTGACGGCGACAAAATTTATTCTGTACTGAAAG  
GTATCGGTACATCTTCAGATGGTCGTTTCAAATCTATTTACGCTCCACGCCCAGATG  
GCCAAGCAAAAGCGCTAAAACGTGCTTATGAAGATGCCGGTTTTGCCCTGAAACAT  
GTGGTCTAATTGAAGGCCATGGTACGGGTACCAAAGCGGGTGATGCCGCAGAATTTG

FIG. 6-1



88/134

CTGGCTTGACCAAACACTTTGGCGCCGCCAGTGATGAAAAGCAATATATCGCCTTAG  
GCTCAGTTAAATCGCAAATTGGTCATACTAAATCTGCGGCTGGCTCTGCGGGTATGA  
TTAAGGCGGCATTAGCGCTGCATCATAAAATCTTACCTGCAACGATCCATATCGATA  
AACCAAGTGAAGCCTTGGATATCAAAAACAGCCCGTTATACCTAAACAGCGAAACGC  
GTCCTTGGATGCCACGTGAAGATGGTATTCCACGTCTGTCAGGTATCAGCTCATTTG  
GTTTTGGCGGCACCAACTTCCATATTATTTTAGAAGAGTATCGCCCAGGTCACGATA  
GCGCATATCGCTTAAACTCAGTGAGCCAACTGTGTTGATCTCGGCAAACGACCAAC  
AAGGTATTGTTGCTGAGTTAAATAACTGGCGTACTAACTGGCTGTCGATGCTGATC  
ATCAAGGGTTTGTATTTAATGAGTTAGTGACAACGTGGCCATTAAAAACCCCATCCG  
TTAACCAAGCTCGTTTAGGTTTTGTTGCGCGTAATGCAAATGAAGCGATCGCGATGA  
TTGATACGGCATTGAAACAATTCAATGCGAACGCAGATAAAATGACATGGTCAGTAC  
CTACCGGGGTTTACTATCGTCAAGCCGGTATTGATGCAACAGGTAAAGTGGTTGCGC  
TATTCTCAGGGCAAGGTTGCAATACGTGAACATGGGTCGTGAATTAACCTGTAAC  
TCCCAAGCATGATGCACAGTGCTGCGGCGATGGATAAAGAGTTCAGTGCCGCTGGTT  
TAGGCCAGTTATCTGCAGTTACTTTCCCTATCCCTGTTTATACGGATGCCGAGCGTA  
AGCTACAAGAAGAGCAATTACGTTTAACGCAACATGCGCAACCAGCGATTGGTAGTT  
TGAGTGTTGGTCTGTTCAAAACGTTTAAGCAAGCAGGTTTTAAAGCTGATTTTGCTG  
CCGGTCATAGTTTCGGTGAGTTAACCGCATTATGGGCTGCCGATGTATTGAGCGAAA  
GCGATTACATGATGTTAGCGCGTAGTCGTGGTCAAGCAATGGCTGCGCCAGAGCAAC  
AAGATTTTGATGCAGGTAAGATGGCCGCTGTTGTTGGTGATCCAAAGCAAGTCGCTG  
TGATCATTGATACCCCTTGATGATGTCTCTAFTGCTAACTTCAACTCGAATAACCAAG  
TTGTTATTGCTGGTACTACGGAGCAGGTTGCTGTAGCGGTTACAACCTTAGGTAATG  
CTGGTTTCAAAGTTGTGCCACTGCCGGTATCTGCTGCGTTCCATACACCTTTAGTTC  
GTCACGCGCAAAAACCATTTGCTAAAGCGGTTGATAGCGCTAAATTTAAAGCGCCAA  
GCATTCCAGTGTTTGCTAATGGCACAGGCTTGGTGATTCAAGCAAACCGAATGACA  
TTAAGAAAAACCTGAAAAACCACATGCTGGAATCTGTTCAATTCAATCAAGAAATTG

FIG. 6-2

89/134

ACAACATCTATGCTGATGGTGGCCGCGTATTTATCGAATTGGTCCAAAGAATGTAT  
TAACTAAATTGGTTGAAAACATTCTCACTGAAAAATCTGATGTGACTGCTATCGCGG  
TTAATGCTAATCCTAAACAACCTGCGGACGTACAAATGCGCCAAGCTGCGCTGCAAA  
TGGCAGTGCTTGGTGTCGCATTAGACAATATTGACCCGTACGACGCCGTTAAGCGTC  
CACTTGTTGCGCCGAAAGCATCACCAATGTTGATGAAGTTATCTGCAGCGTCTTATG  
TTAGTCCGAAAACGAAGAAAGCGTTTGCTGATGCATTGACTGATGGCTGGACTGTTA  
AGCAAGCGAAAGCTGTACCTGCTGTTGTGTCACAACCACAAGTGATTGAAAAGATCG  
TTGAAGTTGAAAAGATAGTTGAACGCATTGTGCAAGTAGAGCGTATTGTGCAAGTAG  
AAAAAATCGTCTACGTTAATGCTGACGGTTCGCTTATATCGCAAATAATCAAGACG  
TTAACAGCGCTGTTGTTAGCAACGTGACTAATAGCTCAGTGACTCATAGCAGTGATG  
CTGACCTTGTTGCCTCTATTGAACGCAGTGTTGGTCAATTTGTTGCACACCAACAGC  
AATTATTAAATGTACATGAACAGTTTATGCAAGGTCCACAAGACTACGCGAAAACAG  
TGCAGAACGTACTTGCTGCGCAGACGAGCAATGAATTACCGGAAAGTTTAGACCGTA  
CATTGTCTATGTATAACGAGTTCCAATCAGAAACGCTACGTGTACATGAAACGTACC  
TGAACAATCAGACGAGCAACATGAACACCATGCTTACTGGTGCTGAAGCTGATGTGC  
TAGCAACCCCAATAACTCAGGTAGTGAATACAGCCGTTGCCACTAGTCACAAGGTAG  
TTGCTCCAGTTATTGCTAATACAGTGACGAATGTTGTATCTAGTGTCAGTAATAACG  
CGGCGGTTGCAGTGCAAACCTGTGGCATTAGCGCCTACGCAAGAAATCGCTCCAACAG  
TCGCTACTACGCCAGCACCCGCATTGGTTGCTATCGTGGCTGAACCTGTGATTGTTG  
CGCATGTTGCTACAGAAGTTGCACCAATTACACCATCAGTTACACCAGTTGTGCGAA  
CTCAAGCGGCTATCGATGTAGCAACTATTAACAAAGTAATGTTAGAAGTTGTTGCTG  
ATAAAACCGGTTATCCAACGGATATGCTGGAACCTGAGCATGGACATGGAAGCTGACT  
TAGGTATCGACTCAATCAAACGTGTTGAGATATTAGGCGCAGTACAGGAATTGATCC  
CTGACTTACCTGAACTTAATCCTGAAGATCTTGCTGAGCTACGCACGCTTGGTGAGA  
TTGTGCGATTACATGAATTCAAAAGCCCAGGCTGTAGCTCCTACAACAGTACCTGTAA

**FIG. 6-3**

90/134

CAAGTGCACCTGTTTCGCCTGCATCTGCTGGTATTGATTTAGCCACATCCAAAACG  
TAATGTTAGAAGTGGTTGCAGACAAAACCGTTACCCAACAGACATGCTAGAACTGA  
GCATGGATATGGAAGCTGACTTAGGTATTGATTCAATCAAGCGTGTGGAAATCTTAG  
GTGCAGTACAGGAGATCATAACTGATTTACCTGAGCTAAACCCTGAAGATCTTGCTG  
AATTACGCACCCTAGGTGAAATCGTTAGTTACATGCAAAGCAAAGCGCCAGTCGCTG  
AAAGTGCGCCAGTGGCGACGGCTCCTGTAGCAACAAGCTCAGCACCGTCTATCGATT  
TGAACCACATTCAAACAGTGATGATGGATGTAGTTGCAGATAAGACTGGTTATCCAA  
CTGACATGCTAGAACTTGGCATGGACATGGAAGCTGATTTAGGTATCGATTCAATCA  
AACGTGTGGAAATATTAGGCGCAGTGCAGGAGATCATCACTGATTTACCTGAGCTAA  
ACCCAGAAGACCTCGCTGAATTACGCACGCTAGGTGAAATCGTTAGTTACATGCAAA  
GCAAAGCGCCAGTCGCTGAGAGTGCGCCAGTAGCGACGGCTTCTGTAGCAACAAGCT  
CTGCACCGTCTATCGATTTAAACCATATCCAAACAGTGATGATGGAAGTGGTTGCAG  
ACAAAACCGGTTATCCAGTAGACATGTTAGAACTTGCTATGGACATGGAAGCTGACC  
TAGGTATCGATTCAATCAAGCGTGTAGAAATTTTAGGTGCGGTACAGGAAATCATT  
CTGACTTACCTGAGCTTAACCCTGAAGATCTTGCTGAACTACGTACATTAGGTGAAA  
TCGTTAGTTACATGCAAAGCAAAGCGCCCGTAGCTGAAGCGCCTGCAGTACCTGTTG  
CAGTAGAAAGTGCACCTACTAGTGTAACAAGCTCAGCACCGTCTATCGATTTAGACC  
ACATCCAAAATGTAATGATGGATGTTGTTGCTGATAAGACTGGTTATCCTGCCAATA  
TGCTTGAATTAGCAATGGACATGGAAGCCGACCTTGGTATTGATTCAATCAAGCGTG  
TTGAAATTCTAGGCGCGGTACAGGAGATCATTACTGATTTACCTGAACTAAACCCAG  
AAGACTTAGCTGAACTACGTACGTTAGAAGAAATTGTAACCTACATGCAAAGCAAGG  
CGAGTGGTGTACTGTAAATGTAGTGGCTAGCCCTGAAAATAATGCTGTATCAGATG  
CATTTATGCAAAGCAATGTGGCGACTATCACAGCGGCCGCAGAACATAAGGCGGAAT  
TTAAACCGGCGCCGAGCGCAACCGTTGCTATCTCTCGTCTAAGCTCTATCAGTAAAA  
TAAGCCAAGATTGTAAAGGTGCTAACGCCCTTAATCGTAGCTGATGGCACTGATAATG  
CTGTGTTACTTGCAGACCACCTATTGCAAACCTGGCTGGAATGTAAGTGCATTGCAAC  
CAACTTGGGTAGCTGTAACAACGACGAAAGCATTTAATAAGTCAGTGAACCTGGTGA

FIG. 6-4

CTTTAAATGGCGTTGATGAACTGAAATCAACAACATTATTACTGCTAACGCACAAT  
TGGATGCAGTTATCTATCTGCACGCAAGTAGCGAAATTAATGCTATCGAATACCCAC  
AAGCATCTAAGCAAGGCCTGATGTTAGCCTTCTTATTAGCGAAATTGAGTAAAGTAA  
CTCAAGCCGCTAAAGTGCGTGGCGCCTTTATGATTGTTACTCAGCAGGGTGGTTCAT  
TAGGTTTTGATGATATCGATTCTGCTACAAGTCATGATGTGAAAACAGACCTAGTAC  
AAAGCGGCTTAAACGGTTTAGTTAAGACACTGTCTCACGAGTGGGATAACGTATTCT  
GTCGTGCGGTTGATATTGCTTCGTCATTAACGGCTGAACAAGTTGCAAGCCTTGTTA  
GTGATGAACTACTTGATGCTAACACTGTATTAACAGAAGTGGGTATCAACAAGCTG  
GTAAAGGCCTTGAACGTATCACGTTAACTGGTGTGGCTACTGACAGCTATGCATTAA  
CAGCTGGCAATAACATCGATGCTAACTCGGTATTTTGTAGTGAGTGGTGGCGCAAAG  
GTGTAAGTGCACATTGTGTTGCTCGTATAGCTAAAGAATATCAGTCTAAGTTCATCT  
TATTGGGACGTTCAACGTTCTCAAGTGACGAACCGAGCTGGGCAAGTGGTATTACTG  
ATGAAGCGGCGTTAAAGAAAGCAGCGATGCAGTCTTTGATTACAGCAGGTGATAAAC  
CAACACCCGTTAAGATCGTACAGCTAATCAAACCAATCCAAGCTAATCGTGAAATTG  
CGCAAACCTTGTCTGCAATTACCGCTGCTGGTGGCCAAGCTGAATATGTTTCTGCAG  
ATGTAAGTAAATGCAGCAAGCGTACAAATGGCAGTCGCTCCAGCTATCGCTAAGTTCG  
GTGCAATCACTGGCATCATTCATGGCGCGGGTGTGTTAGCTGACCAATTCATTGAGC  
AAAAACACTGAGTGATTTTGTAGTCTGTTTACAGCACTAAAATTGACGGTTTGTAT  
CGCTACTATCAGTCACTGAAGCAAGCAACATCAAGCAATTGGTATTGTTCTCGTCAG  
CGGCTGGTTTCTACGGTAACCCCGGCCAGTCTGATTACTCGATTGCCAATGAGATCT  
TAAATAAAACCGCATACCGCTTTAAATCATTGCACCCACAAGCTCAAGTATTGAGCT  
TTAACTGGGGTCTTGGGACGGTGGCATGGTAACGCCTGAGCTTAAACGTATGTTTG  
ACCAACGTGGTGTTTACATTATTCCACTTGATGCAGGTGCACAGTTATTGCTGAATG  
AACTAGCCGCTAATGATAACCGTTGTCCACAAATCCTCGTGGGTAATGACTTATCTA  
AAGATGCTAGCTCTGATCAAAAGTCTGATGAAAAGAGTACTGCTGTAAAAAAGCCAC  
AAGTTAGTCGTTTATCAGATGCTTTAGTAACTAAAAGTATCAAAGCGACTAACAGTA

**FIG. 6-5**

92/134

GCTCTTTATCAAACAAGACTAGTGCTTTATCAGACAGTAGTGCTTTTCAGGTTAACG  
AAAACCACTTTTTAGCTGACCACATGATCAAAGGCAATCAGGTATTACCAACGGTAT  
GCGCGATTGCTTGGATGAGTGATGCAGCAAAAGCGACTTATAGTAACCGAGACTGTG  
CATTGAAGTATGTCGGTTTCGAAGACTATAAATTGTTTAAAGGTGTGGTTTTTGATG  
GCAATGAGGCGGCGGATTACCAAATCCAATTGTCGCCTGTGACAAGGGCGTCAGAAC  
AGGATTCTGAAGTCCGTATTGCCGCAAAGATCTTTAGCCTGAAAAGTGACGGTAAAC  
CTGTGTTTTATTATGCAGCGACAATATTGTTAGCAACTCAGCCACTTAATGCTGTGA  
AGGTAGAACTTCCGACATTGACAGAAAGTGTTGATAGCAACAATAAAGTAACTGATG  
AAGCACAAGCGTTATACAGCAATGGCACCTTGTTCCACGGTGAAAGTCTGCAGGGCA  
TTAAGCAGATATTAAGTTGTGACGACAAGGGCCTGCTATTGGCTTGTCAGATAACCG  
ATGTTGCAACAGCTAAGCAGGGATCCTTCCCGTTAGCTGACAACAATATCTTTGCCA  
ATGATTTGGTTTTATCAGGCTATGTTGGTCTGGGTGCGCAAACAATTTGGTTTAGGTA  
GCTTACCTTCGGTGACAACGGCTTGACTGTGTATCGTGAAGTGGTTGTAGATGAAG  
TATTTTATCTGCAACTTAATGTTGTTGAGCATGATCTATTGGGTTACGCGGCAGTA  
AAGCCCGTTGTGATATTCAATTGATTGCTGCTGATATGCAATTACTTGCCGAAGTGA  
AATCAGCGCAAGTCAGTGTGAGTGACATTTTGAACGATATGTCATGATCGAGTAAAT  
AATAACGATAGGCGTCATGGTGAGCATGGCGTCTGCTTTCTTCATTTTTTAACATTA  
ACAATATTAATAGCTAAACGCGGTTGCTTTAAACCAAGTAAACAAGTGCTTTTAGCT  
ATTACTATTCCAAACAGGATATTAAAGAGAATATGACGGAATTAGCTGTTATTGGTA  
TGGATGCTAAATTTAGCGGACAAGACAATATTGACCGTGTGGAACGCGCTTTCTATG  
AAGGTGCTTATGTAGGTAATGTTAGCCGCGTTAGTACCGAATCTAATGTTATTAGCA  
ATGGCGAAGAACAAGTTATTACTGCCATGACAGTTCTTAACTCTGTCAGTCTACTAG  
CGCAAACGAATCAGTTAAATATAGCTGATATCGCGGTGTTGCTGATTGCTGATGTAA  
AAAGTGCTGATGATCAGCTTGTTAGTCCAAATTGCATCAGCAATTGAAAACAGTGTG  
CGAGTTGTGTTGTTATTGCTGATTTAGGCCAAGCATTAAATCAAGTAGCTGATTTAG

FIG. 6-6

93/134

TTAATAACCAAGACTGTCCTGTGGCTGTAATTGGCATGAATAACTCGGTAAATTTAT  
CTCGTCATGATCTTGAATCTGTAAGTCAACAATCAGCTTTGATGAAACCTTCAATG  
GTTATAACAATGTAGCTGGGTTTCGCGAGTTTACTTATCGCTTCAACTGCGTTTGCCA  
ATGCTAAGCAATGTTATATATACGCCAACATTAAGGGCTTCGCTCAATCGGGCGTAA  
ATGCTCAATTTAACGTTGGAAACATTAGCGATACTGCAAAGACCGCATTCAGCAAG  
CTAGCATAACTGCAGAGCAGGTTGGTTTGTAGAAAGTGTGAGCAGTCGCTGATTCGG  
CAATCGCATTGTCTGAAAGCCAAGGTTTAATGTCTGCTTATCATCATACGCAAACCTT  
TGCATACTGCATTAAGCAGTGCCCGTAGTGTGACTGGTGAAGGCGGGTGTTTTTCAC  
AGGTGCGCAGGTTTATTGAAATGTGTAATTGGTTTACATCAACGTTATATTCCGGCGA  
TTAAAGATTGGCAACAACCGAGTGACAATCAAATGTCACGGTGGCGGAATTCACCAT  
TCTATATGCCTGTAGATGCTCGACCTTGGTCCCACATGCTGATGGCTCTGCACACA  
TTGCCGCTTATAGTTGTGTGACTGCTGACAGCTATTGTCATATTCTTTTACAAGAAA  
ACGCTTACAAGAACTTGTTTTGAAAGAAACAGTCTTGCAAGATAATGACTTAACTG  
AAAGCAAGCTTCAGACTCTTGAACAAAACAATCCAGTAGCTGATCTGCGCACTAATG  
GTTACTTTGCATCGAGCGAGTTAGCATTAAATCATAGTACAAGGTAATGACGAAGCAC  
AATTACGCTGTGAATTAGAACTATTACAGGGCAGTTAAGTACTACTGGCATAAGTA  
CTATCAGTATTAAACAGATCGCAGCAGACTGTTATGCCCGTAATGATACTAACAAAG  
CCTATAGCGCAGTGCTTATTGCCGAGACTGCTGAAGAGTTAAGCAAAGAAATAACCT  
TGGCGTTTGCTGGTATCGCTAGCGTGTTTAAATGAAGATGCTAAAGAATGGAAAACCC  
CGAAGGGCAGTTATTTTACCGCGCAGCCTGCAAATAAACAGGCTGCTAACAGCACAC  
AGAATGGTGTACCTTCATGTACCCAGGTATTGGTGCTACATATGTTGGTTTAGGGC  
GTGATCTATTTTCATCTATTCCACAGATTTATCAGCCTGTAGCGGCTTTAGCCGATG  
ACATTGGCGAAAGTCTAAAAGATACTTTACTTAATCCACGCAGTATTAGTCGTCATA  
GCTTTAAAGAACTCAAGCAGTTGGATCTGGACCTGCGCGGTAAGTTAGCCAATATCG  
CTGAAGCCGGTGTGGGTTTTGCTTGTGTGTTTACCAAGGTATTTGAAGAAGTCTTTG  
CCGTAAAGCTGACTTTGCTACAGGTTATAGCATGGGTGAAGTAAGCATGTATGCAG  
CACTAGGCTGCTGGCAGCAACCGGGATTGATGAGTGCTCGCCTTGCACAATCGAATA

FIG. 6-7

94/134

CCTTTAATCATCAACTTTGCGGCGAGTTAAGAACACTACGTCAGCATTGGGGCATGG  
ATGATGTAGCTAACGGTACGTTTCGAGCAGATCTGGGAAACCTATAACCATTAAGGCAA  
CGATTGAACAGGTCGAAATTGCCTCTGCAGATGAAGATCGTGTGTATTGCACCATTA  
TCAATACACCTGATAGCTTGTGTAGCCGGTTATCCAGAAGCCTGTCAGCGAGTCA  
TTAAGAATTTAGGTGTGCGTGCAATGGCATTGAATATGGCGAACGCAATTCACAGCG  
CGCCAGCTTATGCCGAATACGATCATATGGTTGAGCTATACCATATGGATGTTACTC  
CACGTATTAATACCAAGATGTATTCAAGCTCATGTTATTTACCGATTCCACAACGCA  
GCAAAGCGATTTCCACAGTATTGCTAAATGTTTGTGTGATGTGGTGGATTTCCAC  
GTTTGGTTAATACCTTACATGACAAAGGTGCGCGGGTATTCATTGAAATGGGTCCAG  
GTCGTTTCGTTATGTAGCTGGGTAGATAAGATCTTAGTTAATGGCGATGGCGATAATA  
AAAAGCAAAGCCAACATGTATCTGTTCTGTGAATGCCAAAGGCACCAGTGATGAAC  
TTACTTATATTCGTGCGATTGCTAAGTTAATTAGTCATGGCGTGAATTTGAATTTAG  
ATAGCTTGTTTAACGGGTCAATCCTGGTTAAAGCAGGCCATATAGCAAACACGAACA  
AATAGTCAACATCGATATCTAGCGCTGGTGAGTTATACCTCATTAGTTGAAATATGG  
ATTTAAAGAGAGTAATTATGGAAAATATTGCAGTAGTAGGTATTGCTAATTTGTTCC  
CGGGCTCACAAGCACCGGATCAATTTTGGCAGCAATTGCTTGAACAACAAGATTGCC  
GCAGTAAGGCGACCGCTGTTCAAATGGGCGTTGATCCTGCTAAATATACCGCCAACA  
AAGGTGACACAGATAAATTTTACTGTGTGCACGGCGGTTACATCAGTGATTTCAATT  
TTGATGCTTCAGGTTATCAACTCGATAATGATTATTTAGCCGGTTTAGATGACCTTA  
ATCAATGGGGGCTTTATGTTACGAAACAAGCCCTTACCGATGCGGGTTATTGGGGCA  
GTACTGCACTAGAAAACCTGTGGTGTGATTTTAGGTAATTTGTCATTCCCAACTAAAT  
CATCTAATCAGCTGTTTATGCCTTTGTATCATCAAGTTGTTGATAATGCCTTAAAGG  
CGGTATTACATCCTGATTTTCAATTAACGCATTACACAGCACCGAAAAAACACATG  
CTGACAATGCATTAGTAGCAGGTTATCCAGCTGCATTGATCGCGCAAGCGGCGGGTC  
TTGGTGGTTCACATTTTGCACCTGGATGCGGCTTGCTTCATCTTGTTATAGCGTTA  
AGTTAGCGTGTGATTACCTGCATACGGGTAAAGCCAACATGATGCTTGCTGGTGCGG

**FIG. 6-8**

95/134

TATCTGCAGCAGATCCTATGTTCTGTAATATGGGTTTCTCGATATTCCAAGCTTACC  
CAGCTAACAATGTACATGCCCCGTTTGACCAAAATTCACAAGGTCTATTTGCCGGTG  
AAGGCGCGGGCATGATGGTATTGAAACGTCAAAGTGATGCAGTACGTGATGGTGATC  
ATATTTACGCCATTATTAAAGGCGGCGCATTATCGAATGACGGTAAAGGCGAGTTTG  
TATTAAGCCCGAACACCAAGGGCCAAGTATTAGTATATGAACGTGCTTATGCCGATG  
CAGATGTTGACCCGAGTACAGTTGACTATATTGAATGTCATGCAACGGGCACACCTA  
AGGGTGACAATGTTGAATTGCGTTCGATGGAAACCTTTTTTCAGTCGCGTAAATAACA  
AACCATTACTGGGCTCGGTTAAATCTAACCTTGGTCATTTGTAACTGCCGCTGGTA  
TGCCTGGCATGACCAAAGCTATGTTAGCGCTAGGTAAAGGTCTTATTCCTGCAACGA  
TTAACTTAAAGCAACCACTGCAATCTAAAAACGGTTACTTTACTGGCGAGCAAATGC  
CAACGACGACTGTGTCTTGGCCAACAACCTCCGGGTGCCAAGGCAGATAAACCGCGTA  
CCGCAGGTGTGAGCGTATTTGGTTTTGGTGGCAGCAACGCCCATTTGGTATTACAAC  
AGCCAACGCAAACACTCGAGACTAATTTTAGTGTTGCTAAACCACGTGAGCCTTTGG  
CTATTATTGGTATGGACAGCCATTTTGGTAGTGCCAGTAATTTAGCGCAGTTCAAAA  
CCTTATTAAATAATAATCAAAATACCTTCCGTGAATTACCAGAACAACGCTGGAAAG  
GCATGGAAAGTAACGCTAACGTCATGCAGTCGTTACAATTACGCAAAGCGCCTAAAG  
GCAGTTACGTTGAACAGCTAGATATTGATTTCTTGCGTTTTAAAGTACCGCCTAATG  
AAAAAGATTGCTTGATCCCGCAACAGTTAATGATGATGCAAGTGGCAGACAATGCTG  
CGAAAGACGGAGGTCTAGTTGAAGGTCGTAATGTTGCGGTATTAGTAGCGATGGGCA  
TGGAACTGGAATTACATCAGTATCGTGGTCGCGTTAATCTAACCACCCAAATTGAAG  
ACAGCTTATTACAGCAAGGTATTAACCTGACTGTTGAGCAACGTGAAGAACTGACCA  
ATATTGCTAAAGACGGTGTTGCCTCGGCTGCACAGCTAAATCAGTATACGAGTTTCA  
TTGGTAATATTATGGCGTCACGATTTTCGGCGTTATGGGATTTTTCTGGTCCTGCTA  
TTACCGTATCGGCTGAAGAAAACCTCTGTTTATCGTTGTGTTGAATTAGCTGAAAATC  
TATTTCAAACCAGTGATGTTGAAGCCGTTATTATTGCTGCTGTTGATTTGTCTGGTT  
CAATTGAAAACATTACTTTACGTCAGCACTACGGTCCAGTTAATGAAAAGGGATCTG

**FIG. 6-9**



96/134

TAAGTGAATGTGGTCCGGTTAATGAAAGCAGTTCAGTAACCAACAATATTCTTGATC  
AGCAACAATGGCTGGTGGGTGAAGGCGCAGCGGCTATTGTGCTTAAACCGTCATCGC  
AAGTCACTGCTGAGCAAGTTTATGCGCGTATTGATGCGGTGAGTTTTGCCCTGGTA  
GCAATGCGAAAGCAATTACGATTGCAGCGGATAAAGCATTAACTTGCTGGTATCA  
GTGCTGCTGATGTAGCTAGTGTTGAAGCACATGCAAGTGGTTTTAGTGCCGAAAATA  
ATGCTGAAAAAACCGCGTTACCGACTTTATACCCAAGCGCAAGTATCAGTTCGGTGA  
AAGCCAATATTGGTCATACGTTTAAATGCCTCGGGTATGGCGAGTATTATTAACCG  
CGCTGCTGTTAGATCAGAATACGAGTCAAGATCAGAAAAGCAAACATATTGCTATTA  
ACGGTCTAGGTCTGTGATAACAGCTGCGCGCATCTTATCTTATCGAGTTCAGCGCAAG  
CGCATCAAGTTGCACCAGCGCCTGTATCTGGTATGGCCAAGCAACGCCCACAGTTAG  
TTAAAACCATCAAACCTCGGTGGTCAGTTAATTAGCAACGCGATTGTTAACAGTGCGA  
GTTTCATCTTTACACGCTATTAAAGCGCAGTTTGCCGGTAAGCACTTAAACAAAGTTA  
ACCAGCCAGTGATGATGGATAACCTGAAGCCCCAAGGTATTAGCGCTCATGCAACCA  
ATGAGTATGTGGTGAAGTGGAGCTGCTAACACTCAAGCTTCTAACATTCAAGCATCTC  
ATGTTCAAGCGTCAAGTCATGCACAAGAGATAGCACCAAACCAAGTTCAAAATATGC  
AAGCTACAGCAGCCGCTGTAAGTTCACCCCTTTCTCAACATCAACACACAGCGCAGC  
CCGTAGCGGCACCGAGCGTTGTTGGAGTGACTGTGAAACATAAAGCAAGTAACCAA  
TTCATCAGCAAGCGTCTACGCATAAAGCATTTTTAGAAAGTCGTTTAGCTGCACAGA  
AAAACCTATCGCAACTTGTTGAATTGCAAACCAAGCTGTCAATCCAACTGGTAGTG  
ACAATACATCTAACAATACTGCGTCAACAAGCAATACAGTGCTAACAAATCCTGTAT  
CAGCAACGCCATTAACTTGTGTCTAATGCGCCTGTAGTAGCGACAAACCTAACCA  
GTACAGAAGCAAAAGCGCAAGCAGCTGCTACACAAGCTGGTTTTTCAGATAAAAGGAC  
CTGTTGGTTACAACCTATCCACCGCTGCAGTTAATTGAACGTTATAATAAACCAGAAA  
ACGTGATTTACGATCAAGCTGATTGGTTGAATTCGCTGAAGGTGATATTGGTAAGG  
TATTTGGTGCTGAATACAATATTATTGATGGCTATTCGCGTCGTGTACGTCTGCCAA  
CCTCAGATTACTTGTTAGTAACACGTGTTACTGAACTTGATGCCAAGGTGCATGAAT

**FIG. 6-10**

97/134

ACAAGAAATCATACATGTGTACTGAATATGATGTGCCTGTTGATGCACCGTTCTTAA  
TTGATGGTCAGATCCCTTGGTCTGTTGCCGTGAATCAGGCCAGTGTGATTTGATGT  
TGATTTTCATATATCGGTATTGATTTCCAAGCGAAAGGCGAACGTGTTTACCGTTTAC  
TTGATTGTGAATTAACCTTTCCTTGAAGAGATGGCTTTTGGTGGCGATACTTTACGTT  
ACGAGATCCACATTGATTTCGTATGCACGTAACGGCGAGCAATTATTATTCTTCTTCC  
ATTACGATTGTTACGTAGGGGATAAGAAGGTACTTATCATGCGTAATGGTTGTGCTG  
GTTTCTTTACTGACGAAGAACTTCTGATGGTAAAGGCGTTATTCATAACGACAAAG  
ACAAAGCTGAGTTTAGCAATGCTGTAAATCATCATTACGCCGTTATTACAACATA  
ACCGTGGTCAATACGATTATAACGACATGATGAAGTTGGTTAATGGTGATGTTGCCA  
GTTGTTTTGGTCCGCAATATGATCAAGGTGGCCGTAATCCATCATTGAAATTCTCGT  
CTGAGAAGTTCCTTGATGATTGAACGTATTACCAAGATAGACCCAACCGGTGGTCATT  
GGGGACTAGGCCTGTTAGAAGGTCAGAAAGATTTAGACCCTGAGCATTGGTATTTCC  
CTTGTCACTTTAAAGGTGATCAAGTAATGGCTGGTTCGTTGATGTGCGAAGGTTGTG  
GCCAAATGGCGATGTTCTTCATGCTGTCTCTTGGTATGCATACCAATGTGAACAACG  
CTCGTTTCCAACCACTACCAGGTGAATCACAAACGGTACGTTGTCTGTTGGGCAAGTAC  
TGCCACAGCGCAATACCTTAACTTACCGTATGGAAGTTACTGCGATGGGTATGCATC  
CACAGCCATTCATGAAAGCTAATATTGATATTTTGCTTGACGGTAAAGTGGTTGTTG  
ATTTCAAAAACCTTGAGCGTGATGATCAGCGAACAAGATGAGCATTTCAGATTACCCTG  
TAACACTGCCGAGTAATGTGGCGCTTAAAGCGATTACTGCACCTGTTGCGTCAGTAG  
CACCAGCATCTTCACCCGCTAACAGCGCGGATCTAGACGAACGTGGTGTGAACCGT  
TTAAGTTTCCTGAACGTCCGTTAATGCGTGTTGAGTCAGACTTGTCTGCACCGAAAA  
GCAAAGGTGTGACACCGATTAAAGCATTTTGAAGCGCCTGCTGTTGCTGGTCATCATA  
GAGTGCCTAACCAAGCACCGTTTACACCTTGGCATATGTTTGAGTTTTCGACGGGTA  
ATATTTCTAACTGTTTCGGTCCTGATTTTGATGTTTATGAAGGTCGTATTCACCTC  
GTACACCTTGTGGCGATTTACAAGTTGTTACTCAGGTTGTAGAAGTGCAGGGCGAAC  
GTCTTGATCTTAAAAATCCATCAAGCTGTGTAGCTGAATACTATGTACCGGAAGACG

**FIG. 6-11**

98/134

CTTGGTACTTTACTAAAAACAGCCATGAAACTGGATGCCTTATTCATTAATCATGG  
AAATTGCATTGCAACCAAATGGCTTTATTTCTGGTTACATGGGCACGACGCTTAAAT  
ACCCTGAAAAAGATCTGTTCTTCCGTAACCTTGATGGTAGCGGCACGTTATTAAAGC  
AGATTGATTTACGCGGCAAGACCATTGTGAATAAATCAGTCTTGGTTAGTACGGCTA  
TTGCTGGTGGCGCGATTATTCAAAGTTTCACGTTTGATATGTCTGTAGATGGCGAGC  
TATTTTATACTGGTAAAGCTGTATTTGGTTACTTTAGTGGTGAATCACTGACTAACC  
AACTGGGCATTGATAACGGTAAAACGACTAATGCGTGGTTTGTTGATAACAATACCC  
CCGCAGCGAATATTGATGTGTTTGATTAACTAATCAGTCATTGGCTCTGTATAAAG  
CGCCTGTGGATAAACCGCATTATAAATTGGCTGGTGGTCAGATGAACTTTATCGATA  
CAGTGTCACTGGTTGAAGGCGGTGGTAAAGCGGGCGTGGCTTATGTTTATGGCGAAC  
GTACGATTGATGCTGATGATTGGTTCTTCCGTTATCACTTCCACCAAGATCCGGTGA  
TGCCAGGTTTATTAGGTGTTGAAGCTATTATTGAGTTGATGCAGACCTATGCGCTTA  
A-~~AA~~ATGATTTGGGTGGCAAGTTTGCTAACCACGTTTCATTGCGCCGATGACGCAAG  
TTGATTGGAAATACCGTGGGCAAATTACGCCGCTGAATAAACAGATGTCACTGGACG  
TGCATATCACTGAGATCGTGAATGACGCTGGTGAAGTGCGAATCGTTGGTGATGCGA  
ATCTGTCTAAAGATGGTCTGCGTATTTATGAAGTTAAAAACATCGTTTTAAGTATTG  
TTGAAGCGTAAAGGGTCAAGTGTAACGTGCTTAAGCGCCGCATTGGTTAAAGACGCT  
TTGCACGCCGTGAATCCGTCCATGGAGGCTTGGGGTTGGCATCCATGCCAACACAG  
CAAGCTTACTTTAATCAATACGGCTTGGTGTCCATTTAGACGCCTCGAACTTAGTAG  
TTAATAGACAAAATAATTTAGCTGTGGAATGAATATAGTAAGTAATCATTTCGGCAGC  
TACAAAAAAGGAATTAAGAATGTCGAGTTTAGGTTTTACAATAACAACGCAATTAA  
CTGGGCTTGGAAGTAGATCCAGCGTCAGTTCATACACAAGATGCAGAAATTAAAGC  
AGCTTTAATGGATCTAACTAAACCTCTCTATGTGGCGAATAATTAGGCGTAACCTGG  
TATAGCTAATCATACGTCAGTAGCAGGTGCGATCAGCAATAACATCGATGTTGATGT  
ATTGGCGTTTGCGCAAAAGTTAAACCCAGAAGATCTGGGTGATGATGCTTACAAGAA  
ACAGCACGGCGTTAAATATGCTTATCATGGCGGTGCGATGGCAAATGGTATTGCCTC

FIG. 6-12

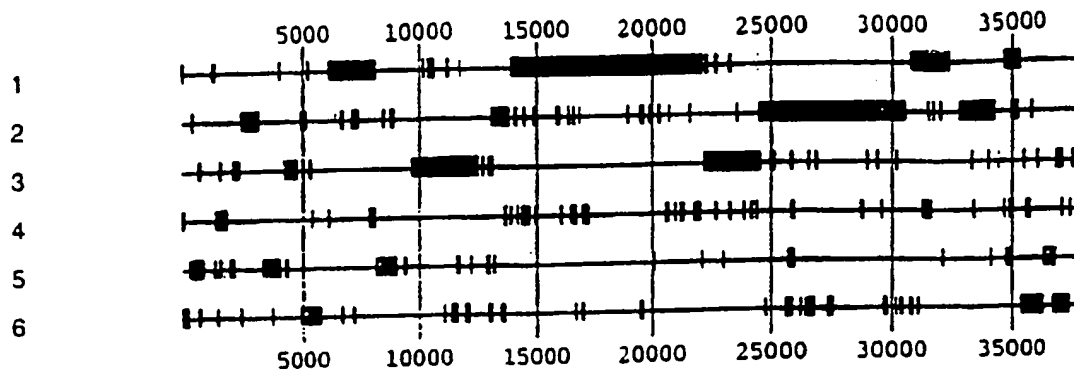
99/134

GGTTGAATTGGTTGTTGCGTTAGGTAAAGCAGGGCTGTTATGTTTCATTTGGTGCTGC  
AGGTCTAGTGCCTGATGCGGTTGAAGATGCAATTCGTCGTATTCAAGCTGAATTACC  
AAATGGCCCTTATGCGGTTAACTTGATCCATGCACCAGCAGAAGAAGCATTAGAGCG  
TGGCGCGGTTGAACGTTTCCTAAACTTGGCGTCAAGACGGTAGAGGCTTCAGCTTA  
CCTTGGTTTAACTGAACACATTGTTTGGTATCGTGCTGCTGGTCTAACTAAAAACGC  
AGATGGCAGTGTTAATATCGGTAACAAGGTTATCGCTAAAGTATCGCGTACCGAAGT  
TGGTCGCCGCTTTATGGAACCTGCACCGCAAAAATTACTGGATAAGTTATTAGAACA  
AAATAAGATCACCCCTGAACAAGCTGCTTTAGCGTTGCTTGTACCTATGGCTGATGA  
TATTACTGGGGAAGCGGATTCTGGTGGTCATACAGATAACCGTCCGTTTTTAACATT  
ATTACCGACGATTATTGGTCTGCGTGATGAAGTGCAAGCGAAGTATAACTTCTCTCC  
TGCATTACGTGTTGGTGCTGGTGGTGGTATCGGAACGCCTGAAGCAGCACTCGCTGC  
ATTTAACATGGGCGCGGCTTATATCGTTCTGGGTTCTGTGAATCAGGCGTGTTGA  
AGCGGGTGCACTCTGAATATACTCGTAACTGTTATCGACAGTTGAAATGGCTGATGT  
GACTATGGCACCTGCTGCAGATATGTTTGAAATGGGTGTGAAGCTGCAAGTATTAAA  
ACGCGGTTCTATGTTGCGGATGCGTGCGAAGAACTGTATGACTTGATGTGGCTTA  
TGACTCGATTGAAGATATCCCAGCTGCTGAACGTGAGAAGATTGAAAAACAAATCTT  
CCGTGCAAAACCTAGACGAGATTTGGGATGGCACTATCGCTTCTTTACTGAACGCGA  
TCCAGAAATGCTAGCCCGTGCAACGAGTAGTCCTAAACGTAAAATGGCACTTATCTT  
CCGTGTTGTTATCTTGGCCTTTCTTCACGCTGGTCAAACACAGGCGAGAAGGGACGTGA  
AATGGATTATCAGATTTGGGCAGGCCCAAGTTTAGGTGCATTCAACAGCTGGGTGAA  
AGGTTCTTACCTTGAAGACTATACCCGCCGTGGCGCTGTAGATGTTGCTTTGCATAT  
GCTTAAAGGTGCTGCGTATTTACAACGTGTAAACCAGTTGAAATTGCAAGGTGTTAG  
CTTAAGTACAGAATTGGCAAGTTATCGTACGAGTGATTAATGTTACTTGATGATATG  
TGAATTAATTAAAGCGCCTGAGGGCGCTTTTTTTGGTTTTTAACTCAGGTGTTGTAA  
CTCGAAATTGCCCCTTTC

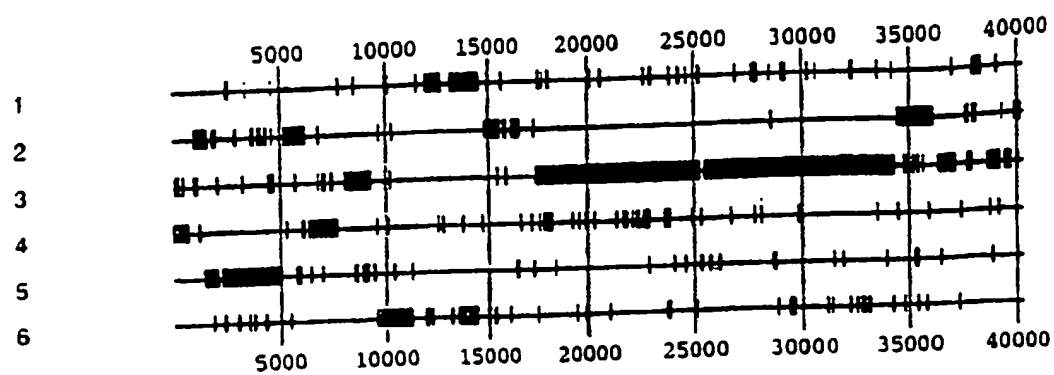
\*  
19227

FIG. 6-13

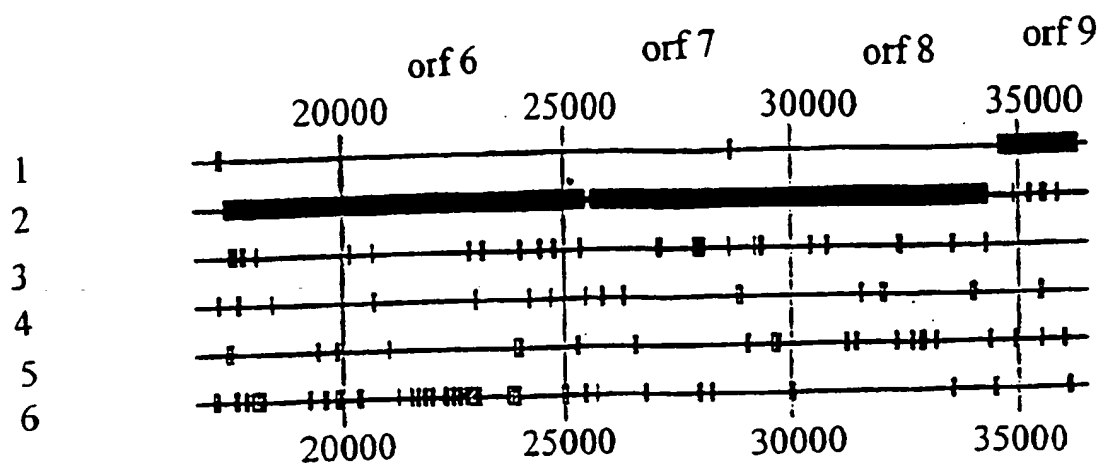
100/134

**FIG. 7A**

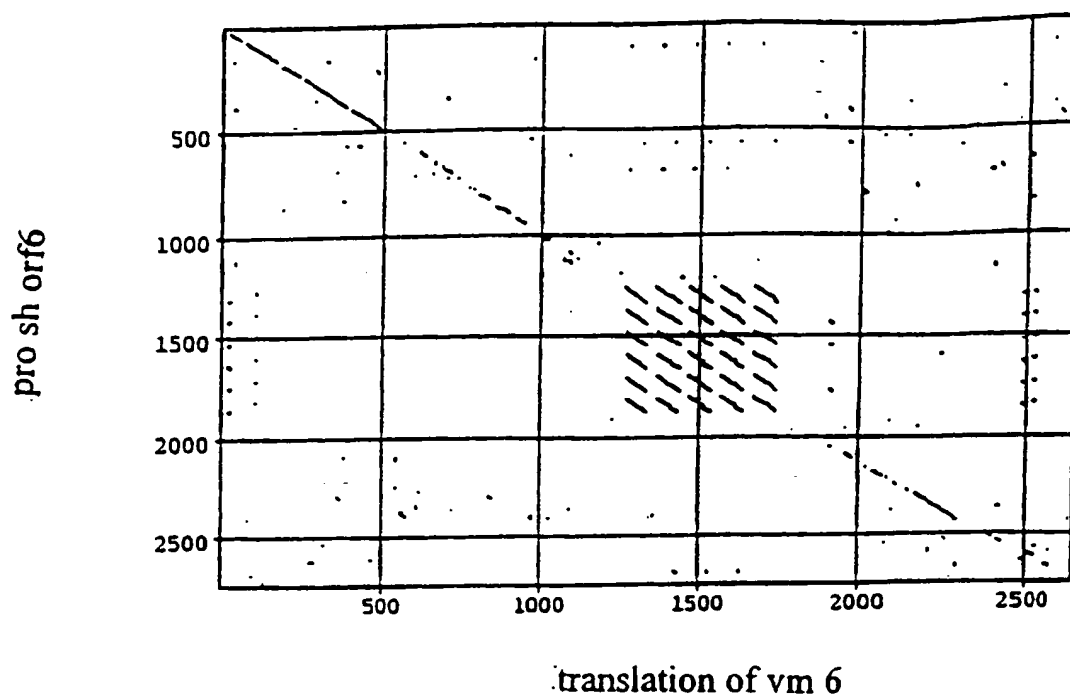
101/134

**FIG. 7B**

102/134

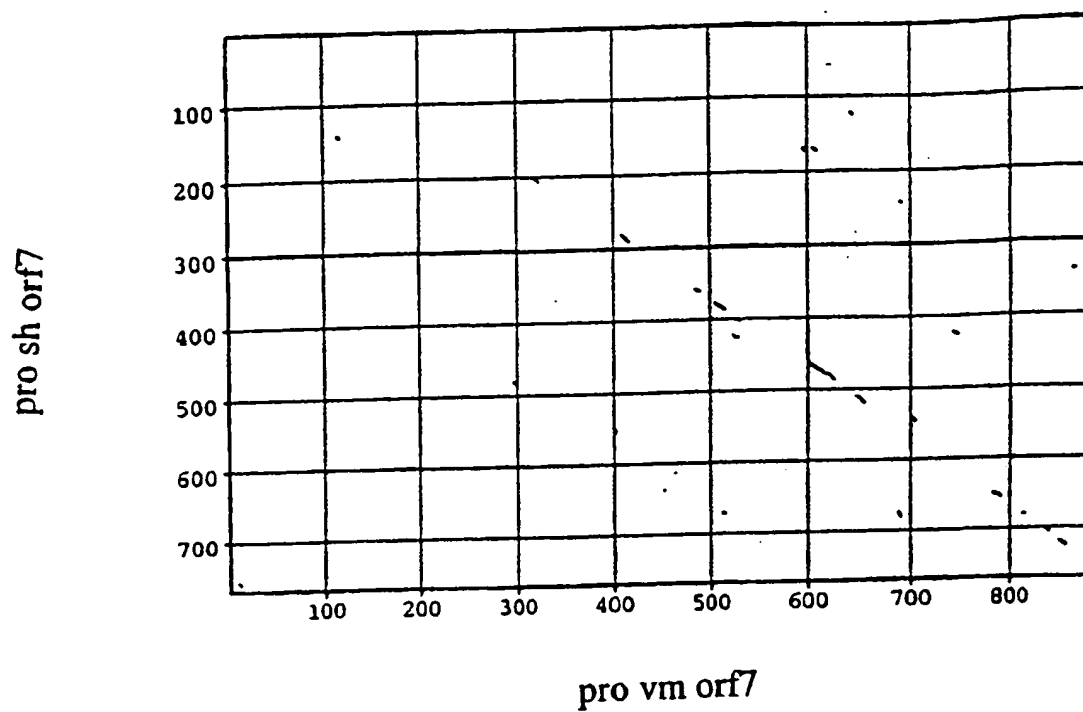
**FIG. 8**

103/134

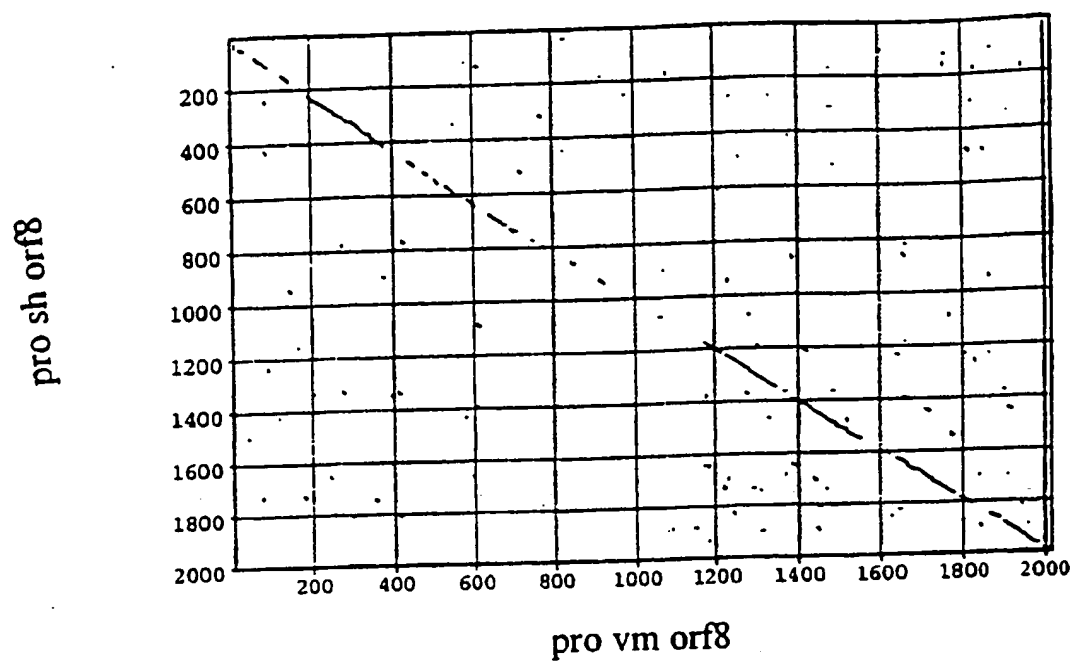
**FIG. 9**



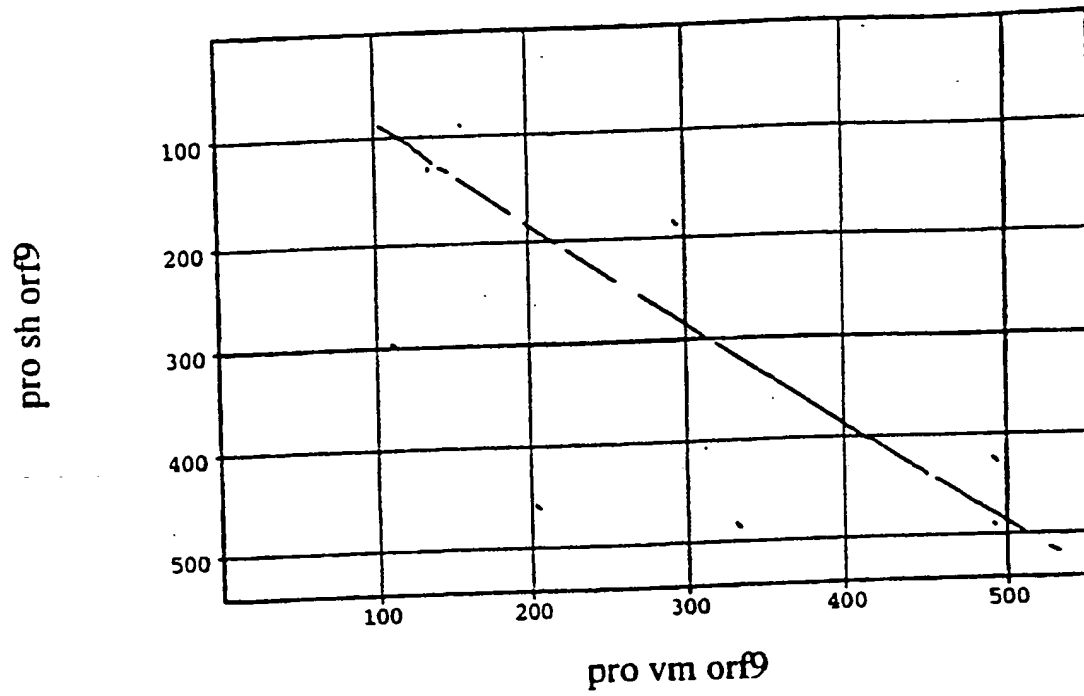
104/134

**FIG. 10**

105/134

**FIG. 11**

106/134

**FIG. 12**

107/134

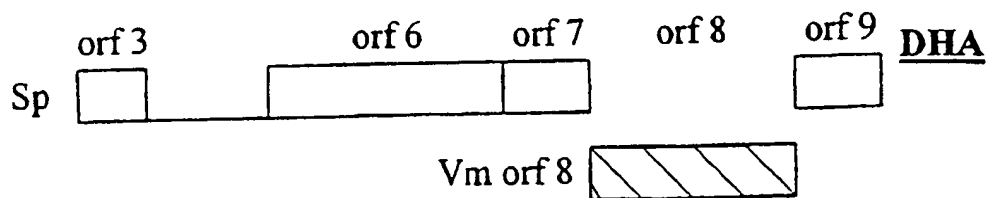
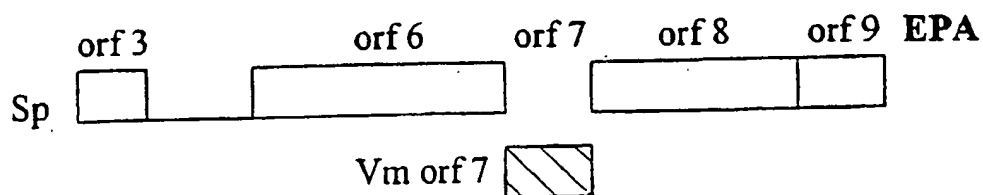
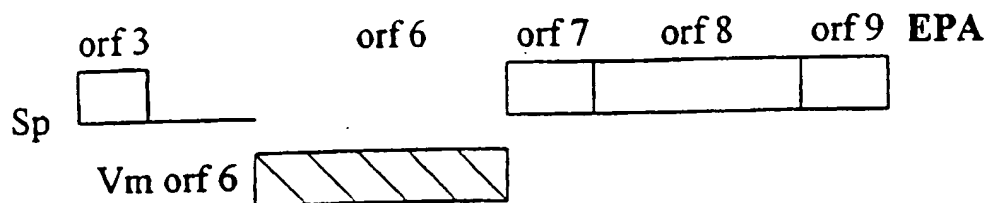
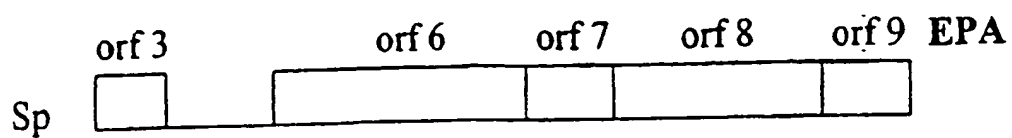


FIG. 13

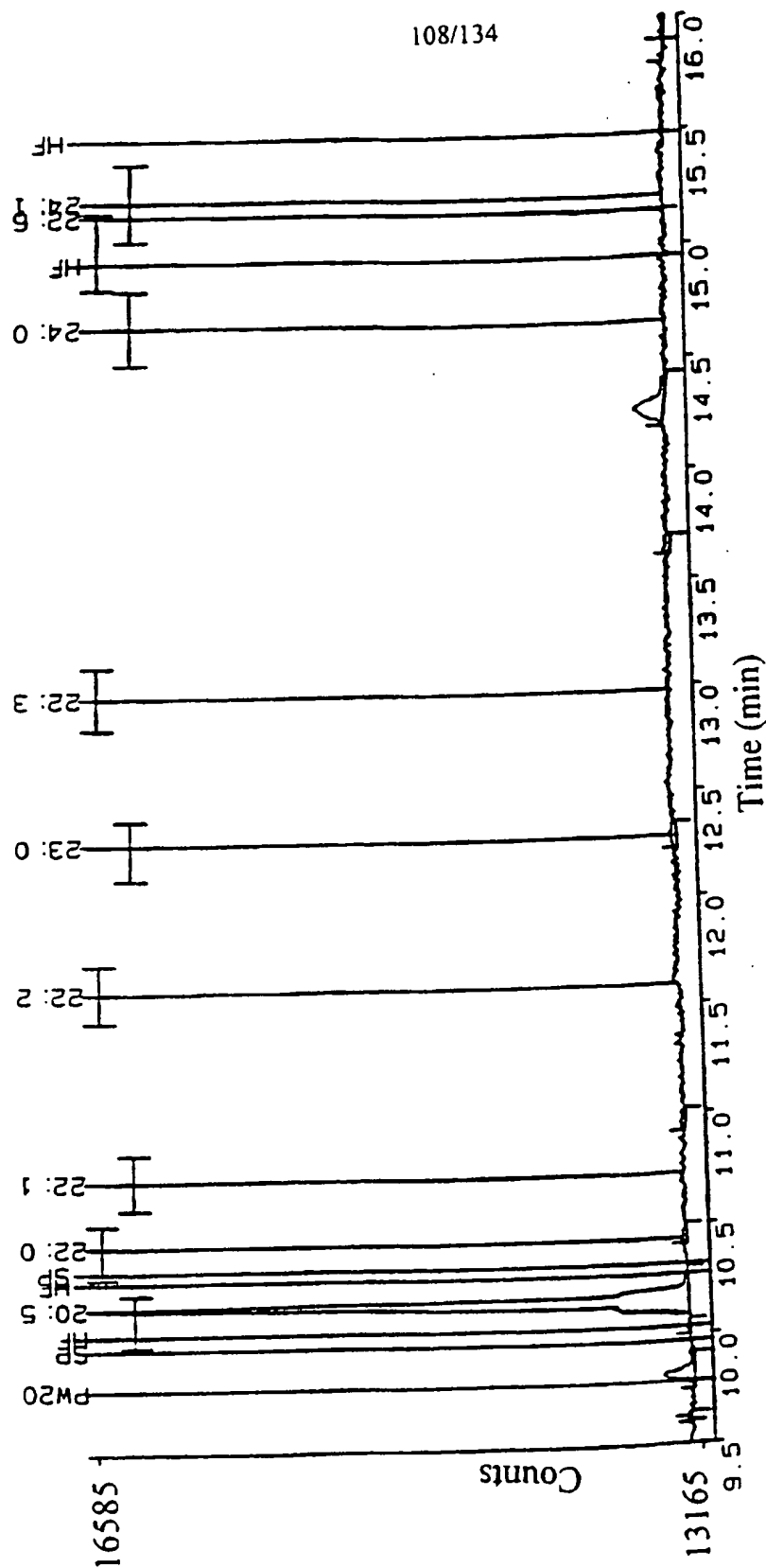
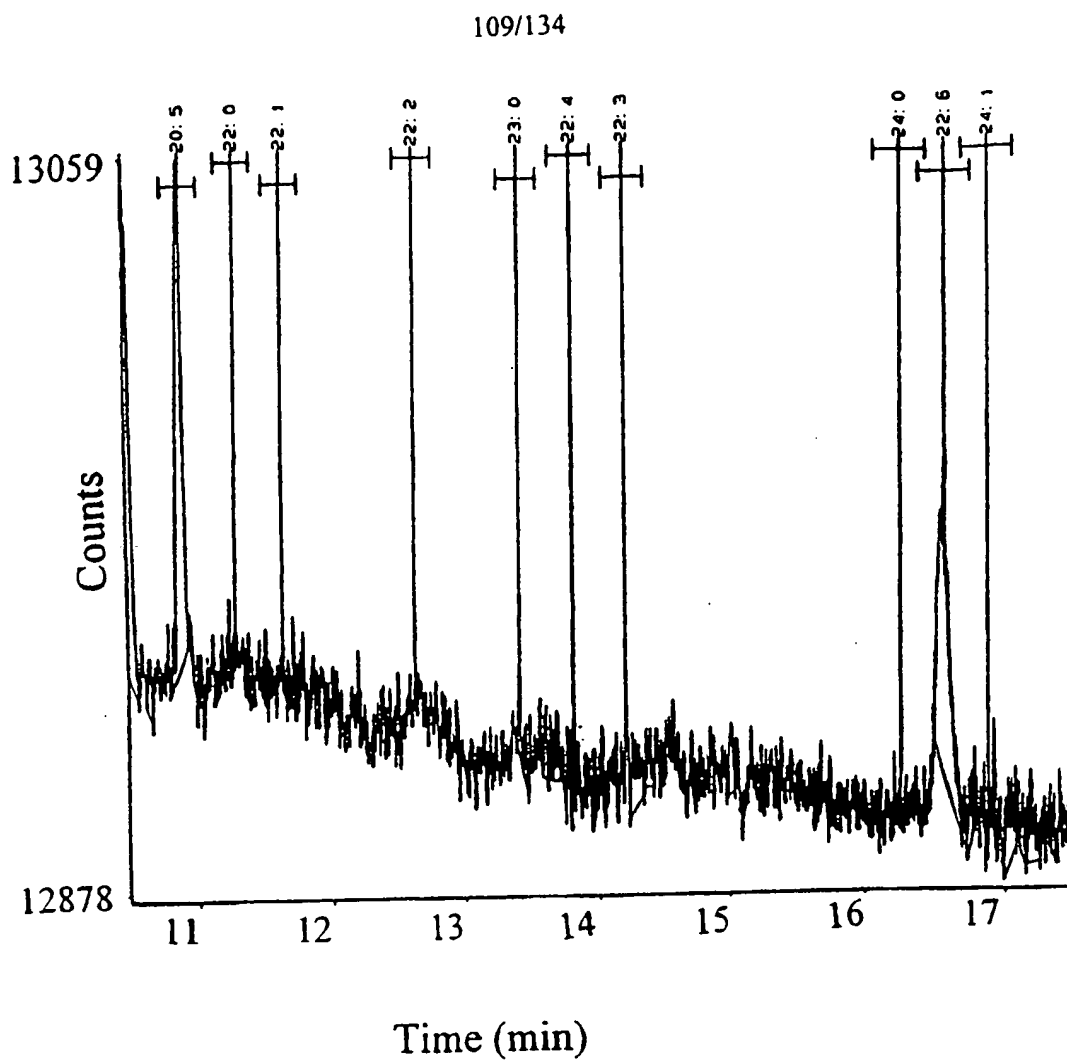


FIG. 14

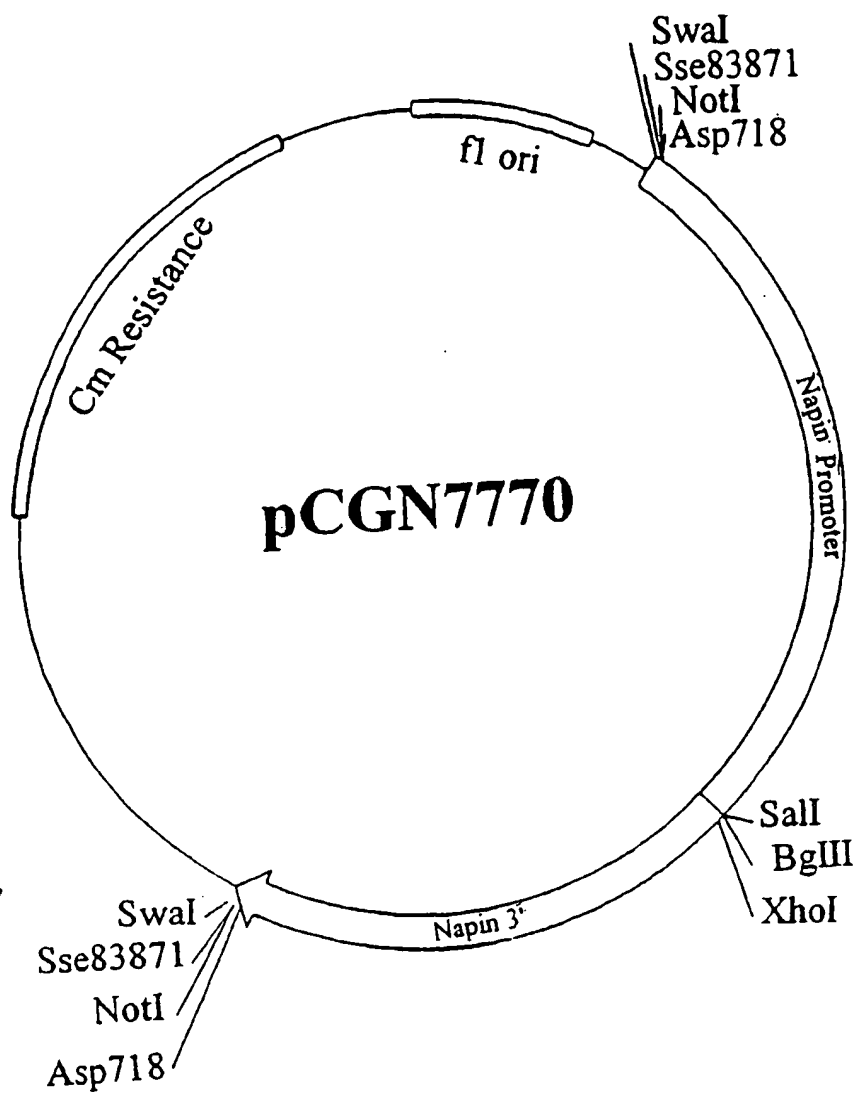
**FIG. 15**

110/134

<u>EPA (%Fatty acids)</u>	<u>DHA (%Fatty acids)</u>	<u>20 deg C</u>
0.00	0.06	pEPAD8
0.60	0.70	4
0.64	0.66	5
0.33	0.22	6s
0.45	0.59	6l
		<u>23 deg C</u>
		pEPAD8
0.02	0.06	4
0.32	0.62	6s
0.27	0.22	6l
0.18	0.65	

FIGURE 16

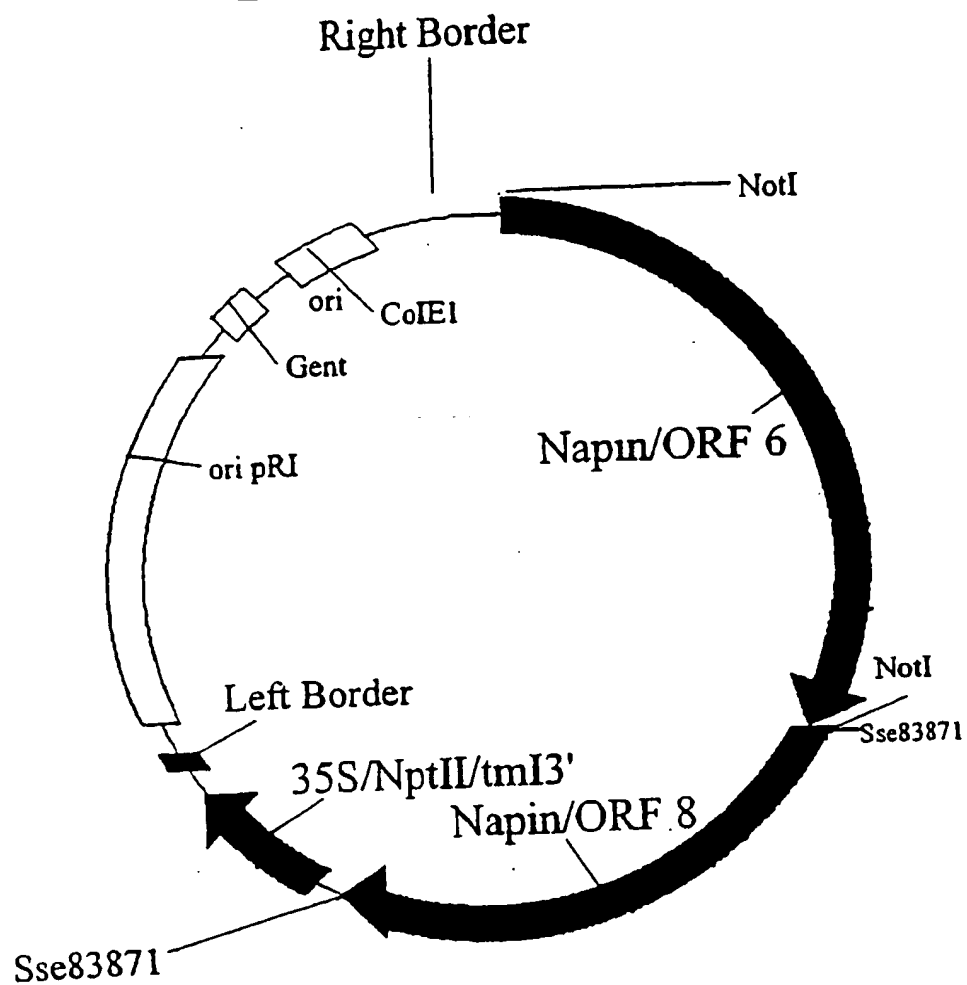
111/134

**FIG. 17**



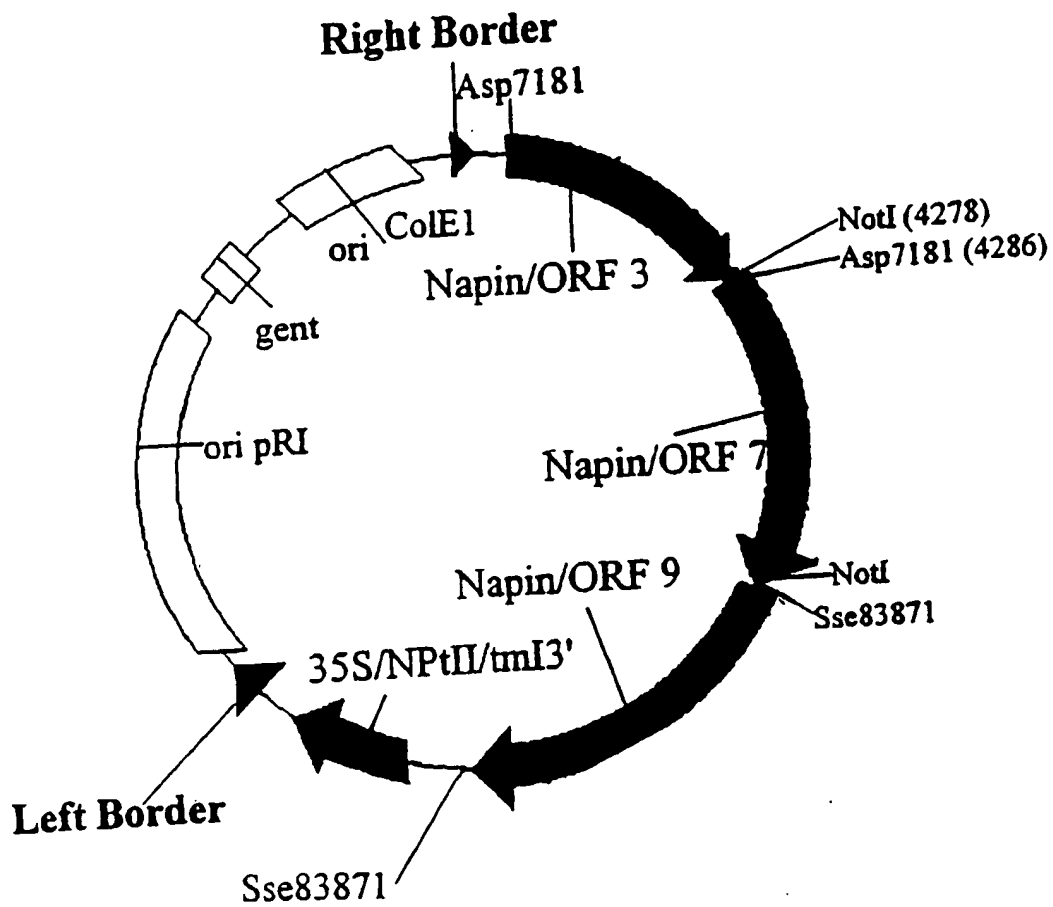
112/134

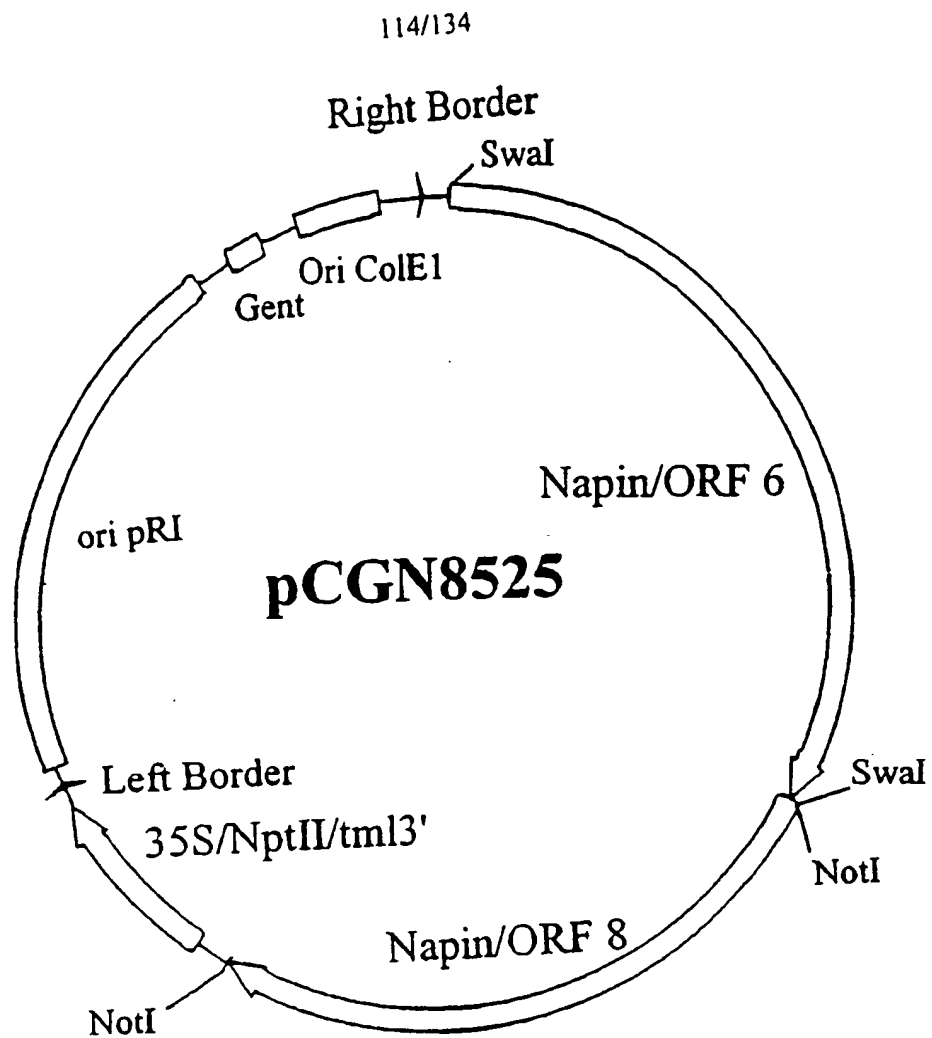
# pCGN8535

**FIG. 18**

113/134

# pCGN8537

**FIG. 19**



**FIG. 20**

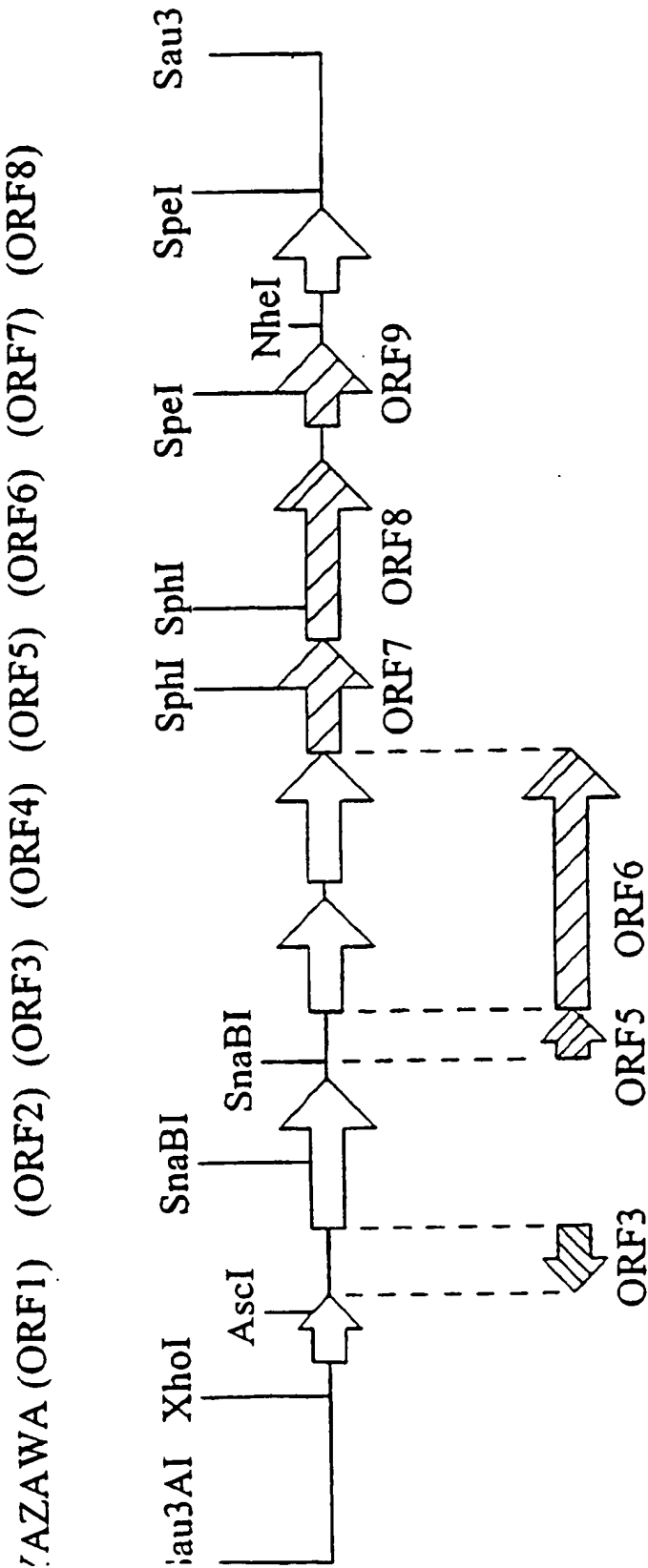
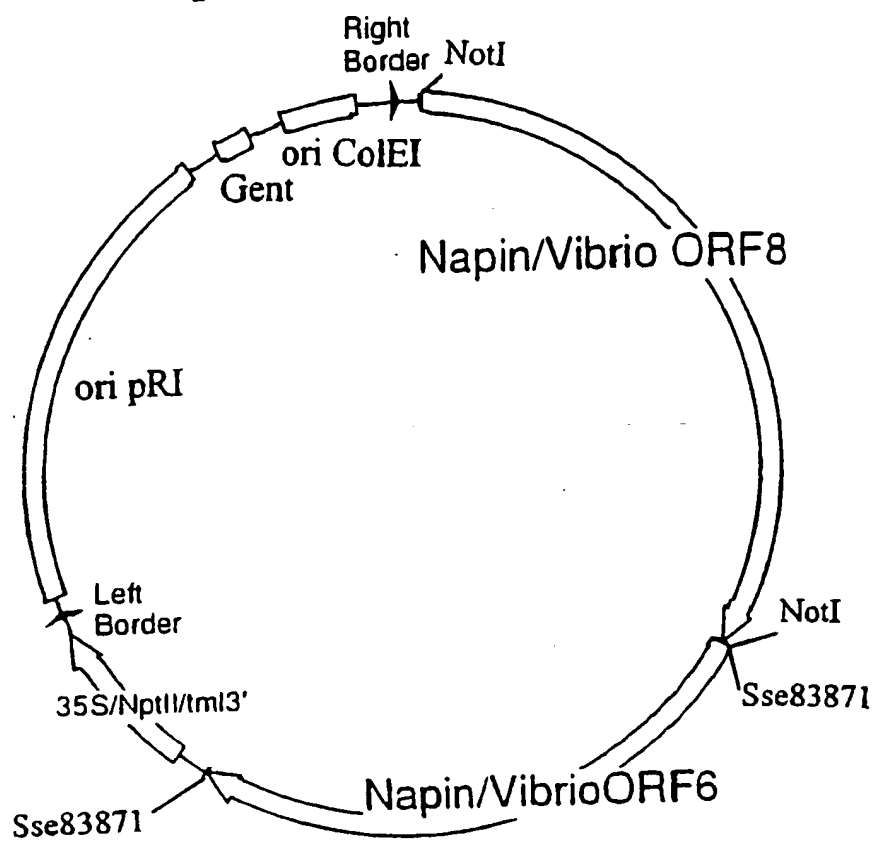


FIG. 21

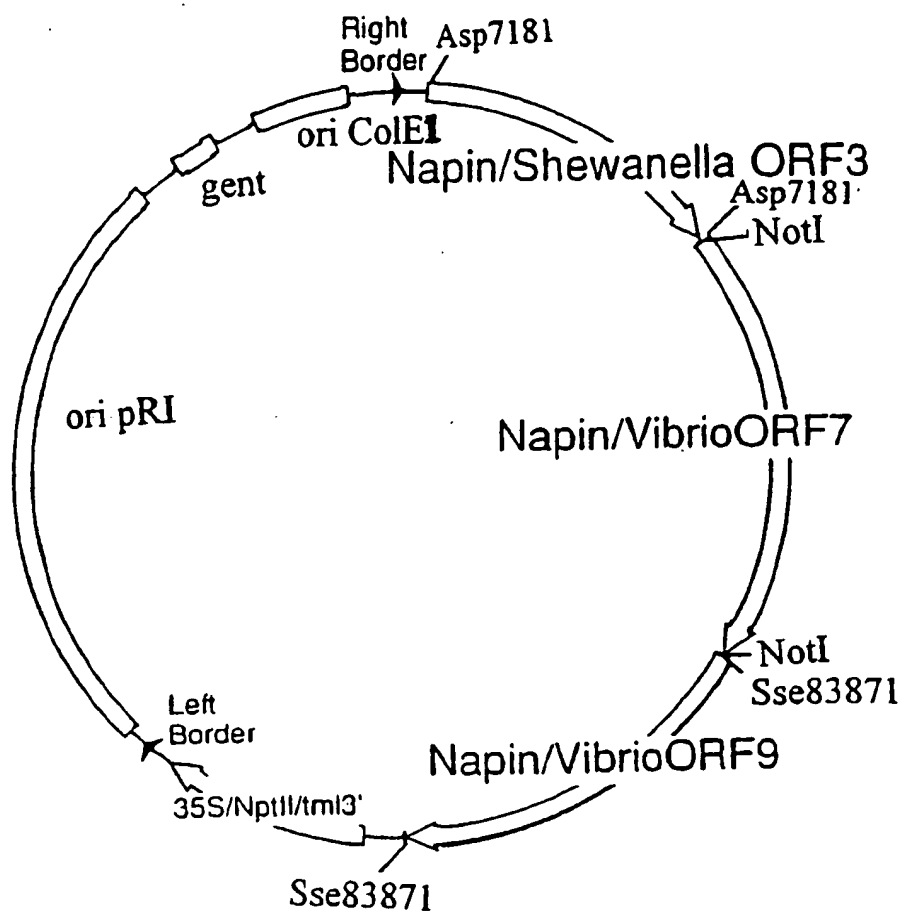
116/134

# pCGN8560

**FIG. 22**

117/134

# pCGN8556

**FIG. 23**

118/134

↓  
ATT GGT AAA AAT AGG GGT TAT GTT TGT TGC TTT AAA GAG TGT CCT GAA  
I G K N R G Y V C C F K E C P E

↓ 9157 ↓ ↓  
AAA TTG CTA ACT TCT CGA TTG ATT TCC TTA TAC TTC TGT CCG TTA ACA  
K L L T S R L I S L Y F C P L T

↓  
ATA CAA GAG TGC GAT AAC CAG ACT ACA GAG TTG GTT AAG TCA TGG CTG  
I Q E C D N Q T T E L V K S W L

↓ ↓  
CCT GAA GAT GAG TTA ATT AAG GTT AAT CGC TAC ATT AAA CAA GAA GCT  
P E D E L I K V N R Y I K Q E A

9016 ↓  
AAA ACT CAA GGT TTA ATG GTA AGA G  
K T Q G L M V R

FIG. 24

119/134

AGCGAAATGC TTATCAAGAA ATTCCAAGAT CAATACATCA CTGGGAAGAA AATTCATTCC 60  
CTGGTTCACT GGGTAACGTT ATTTCCGGCC GTATTGCTAA CCGCTTCGAC CTTGGTGGCA 120  
TGAACGTGTG CGTTGATGCA GCATGTGCAG GCCCTCTTGC TGCATTGCGT ATGGCATTAA 180  
GCGAGCTTGT TGAAGGCCCG AGCGAAATGA TGATTACAGG TGGTGTGTGT ACCGATAACT 240  
CACCAACCAT GTACATGAGC TTCTCTAAAA CACCGGCATT CACGACAAAC GAAACAATTC 300  
AACCATTGCA TATTGACTCG AAAGGTATGA TGATTGGTGA AGGTATCGGT ATGATTGCGC 360  
TTAAACGTCT TGAAGACGCA GAGCGTGATG GCGACCGTAT CTATTCCGTG ATTAAAGGTG 420  
TTGGGTGCAT CTTCAGACGG TAATTATTA AGAGTANTTA TGGCNCNTCGT CCTGAAGGTC 480  
AGGCTAAGGC ACTTAAACGT GCTTACGACG ATGCAGGTTT CGCACCCGCAC ACACCTGGCT 540  
TACTTGAAGC CCACGGCACA GGCACAGCAG CAGGTGATGT GGCAGAATTC AGTGGTCTTA 600  
ACTCTGTATT CAGTGAAGGC AATGACGAAA AGCAACACAT CGCATTAGGT TCAGTGAAT 660  
CACAGATTGG TCACACTAAA TCAACAGCGG GTACTGGGGG TCTAATCAAA GCGTCTTTAG 720  
CACTGCACCA TAAAGTACTG CCGCCCAACAA TCAATGTAAC CAGCCCCTAAC CCTAAACTGA 780  
ATATTGAAGA CTCGCCCTTC TACCTCAATA CACAGACGCG TCCATGGATG CAACGTGTG 840  
ATGGTACACC GCGTCGTGCT GGTATTAGCT CATTTGGTTT TGGTG 885

FIG. 25



120/134

80	*		*	100	*	120	*
CCCCCTTCGTA	CCCCATCCCA	AACGGGTGAT	GCCGCAGAA	TTCGTGGCTT			

impl str +  
TCTAATTGAA GGCCATGGTA C  
| | | | | | | | | | | | | | | |  
TCTAATTGAA GGCCATGGTA C

```

impl str +
3-2 (-VECTO
AGA ACGCAAAGTT GCGCACTGT TTGGTCGCCA
| | | | | | | | | | | | | |
CAA AGCGGGTGAT GCGCACTGT TTGGTCGCTT

```

**FIG. 26-1**



122/134

```

      jmp1 st +
      3-2 (-VECTO

      GCACTGCT GCAAGCATGA ACGCGTCGTT
      || ||| | | ||| ||| ||| |||
      GCTCTGGG GCTATCATTA ACGCGGCATT

      260 * 280 * 300 *
      * * *
      3-2 (-VECTO  CATAAAATCT TACCTGCAAC GATCCATATC GATAAACCAA GTGAAGCCTT
      *
      ACGCGTGCAT
      jmp1 st +
      AACGGTG
      | | |
      3-2 (-VECTO  ACGCGTG

      jmp1 st +
      T
      |
      3-2 (-VECTO  A

      jmp1 st +
      3-2 (-VECTO

      TCCCTGGTGC TAACCATATC AGCAAACCA
      | ||| | ||| ||| ||| |||
      TACCTGCAAC GATCCATATC GATAAACCA

      320 * 340 * 360 *
      * * *
      3-2 (-VECTO  AACAGCCCGT TATACCTAAA CAGCGAAACG CGTCCTTGGA TGCCACGTGA
      *
      GGATATCAAA
  
```

FIG. 26-3

123/134

```

      jmp1 str +
      3-2(-VECTO

      CTCACCTT TGTATCTAAA CACTGAGACT TCGTCCATGG TTACCACGGTGT
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      CAGCCCGT TATACCTAAA CAGCGAAACG GCGTCCTTGG ATGCCACGTGA

      380          400
      *          *
      AGATGGTATT CCACGTCGTG CAGGTATTAG CTCATTGGT TTTGGTGGC

      jmp1 str +
      3-2(-VECTO

      TGATGGTACG CCGCGCCGCG CGGGTATTAG CTCATTGGT TTTGGTGGC>
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      AGATGGTATT CCACGTCGTG CAGGTATTAG CTCATTGGT TTTGGTGGC
  
```

FIG. 26-4

124/134

CGCTGCCGCCGCGTCTCGCCGCGCCGCGCCGCGCCGCGCCGCGCTCGCGCGCACGCC  
CGCGCGTCTCGCCGCGCCTGCTGTCTCGAACGAGCTTCTCGAGAAGGCCGAGACCGTCTG  
TCATGGAGGTCTCGCCGCCAAGACTGGCTACGAGACTGACATGATCGAGTCCGACATG  
GAGCTCGAGACTGAGCTCGGCATTGACTCCATCAAGCGTGTGAGATCCTCTCCGAGGT  
TCAGGCCATGCTCAACGTCGAGGCCAAGGACGTCGACGCTCTCAGCCGCACTCGCACTG  
TGGGTGAGGTGCTCAACGCCATGAAGGCTGAGATCGCTGGTGGCTCTGCCCCGGCGCCT  
GCCGCCGCTGCCCCAGGTCCGGCTGCTGCCGCCCTGCGCCTGCTGTCTCGAGCGAGCT  
TCTCGAGAAGGCCGAGACTGTCGTATGGAGGTCTCGCCGCCAAGACTGGCTACGAGA  
CTGACATGATTGAGTCCGACATGGAGCTCGAGACCGAGCTCGGCATTGACTCCATCAAG  
CGTGTGAGATTCTCTCCGAGGTTGAGGCCATGCTCAACGTCGAGGCCAAGGACGTCGA  
CGCTCTCAGCCGCACTCGCACTGTTGGTGAGGTGCTCGATGCCATGAAGGCTGAGATCG  
CTGGCAGCTCCGCCTCGGCGCCTGCCGCCGCTGCTCCTGCTCCGGCTGCTGCCGCTCCT  
GCGCCCGCTGCCGCCGCCCTGCTGTCTCGAACGAGCTTCTCGAGAAAGCCGAGACTGT  
CGTCATGGAGGTCTCGCCGCCAAGACTGGCTACGAGACTGACATGATCGAGTCCGACA  
TGGAGCTCGAGACTGAGCTCGGCATTGACTCCATCAAGCGTGTGAGATCCTCTCCGAG  
GTTGAGGCCATGCTCAACGTCGAGGCCAAGGACGTCGATGCCCTCAGCCGCACCCGCAC  
TGTTGGCGAGGTTGTCGATGCCATGAAGGCCGAGATCGCTGGTGGCTCTGCCCCGGCGC  
CTGCCGCCGCTGCCCTGCTCCGGCTGCCGCCGCCCTGCTGTCTCGAACGAGCTTCTT  
GAGAAGGCCGAGACTGTCGTATGGAGGTCTCGCCGCCAAGACTGGCTACGAGACCGA  
CATGATCGAGTCCGACATGGAGCTCGAGACCGAGCTCGGCATTGACTCCATCAAGCGTG  
TCGAGATTCTCTCCGAGGTTGAGGCCATGCTCAACGTCGAGGCCAAGGACGTCGATGCT  
CTCAGCCGCACTCGCACTGTTGGCGAGGTGCTCGATGCCATGAAGGCTGAGATCGCCGG  
CAGCTCCGCCCGGCGCCTGCCGCCGCTGCTCCTGCTCCGGCTGCTGCCGCTCCTGCGC  
CCGCTGCCGCTGCCCTGCTGTCTCGAGCGAGCTTCTCGAGAAGGCCGAGACCGTCGTC  
ATGGAGGTCTCGCCGCCAAGACTGGCTACGAGACTGACATGATTGAGTCCGACATGGA  
GCTCGAGACTGAGCTCGGCATTGACTCCATCAAGCGTGTGAGATCCTCTCCGAGGTTCT  
AGGCCATGCTCAACGTCGAGGCCAAGGACGTCGATGCCCTCAGCCGCACCCGCACTGTT  
GGCGAGGTTGTCGATGCCATGAAGGCCGAGATCGCTGGTGGCTCTGCCCCGGCGCCTGC  
CGCCGCTGCCCTGCTCCGGCTGCCGCCGCCCTGCTGTCTCGAACGAGCTTCTTGAGA  
AGGCCGAGACCGTCGTCATGGAGGTCTCGCCGCCAAGACTGGCTACGAGACCGACATG  
ATCGAGTCCGACATGGAGCTCGAGACCGAGCTCGGCATTGACTCCATCAAGCGTGTGGA  
GATTCTCTCCGAGGTTGAGGCCATGCTCAACGTCGAGGCCAAGGACGTCGACGCTCTCA  
GCCGCACTCGCACTGTTGGCGAGGTGCTCGATGCCATGAAGGCTGAGATCGCTGGTGGC  
TCTGCCCGGCGCCTGCCGCCGCTGCTCCTGCCTCGGCTGGCGCCGCGCCTGCGGTCAA  
GATTGACTCGGTCCACGGCGCTGACTGTGATGATCTTCCCTGATGCACGCCAAGGTGG  
TTGACATCCGCCGCCCGGACGAGCTCATCCTGGAGCGCCCCGAGAACC GCCCCGTTCTC  
GTTGTGATGACGGCAGCGAGCTCACCTCGCCCTGGTCCGCGTCTCGGCGCCTGCGC  
CGTTGTCTGACCTTTGAGGGTCTCCAGCTCGCTCAGCGCGCTGGTGCCGCTGCCATCC  
GCCACGTGCTGCCAAGGATCTTCCGCGGAGAGCGCCGAGAAGGCCATCAAGGAGGCC  
GAGCAGCGCTTTGGCGCTCTCGGCGGCTTCATCTCGCAGCAGGCGGAGCGCTTCGAGCC  
CGCCGAAATCCTCGGCTTACGCTCATGTGCGCCAAGTTCGCCAAGGCTTCCCTCTGCA  
CGGCTGTGGCTGGCGGCCGCCCGGCTTTATCGGTGTGGCGCGCCTTGACGGCCGCTC

Figure 27 A-1

125/134

GGATTCACTTCGCGAGGGCACTTCTGACGCGCTCAAGCGTGCCCAGCGTGGTGCCATCTT  
TGGCCTCTGCAAGACCATCGGCCTCGAGTGGTCCGAGTCTGACGTCTTTTCCCGCGGCG  
TGGACATTGCTCAGGGCATGCACCCCGAGGATGCCGCCGTGGCGATTGTGCGCGAGATG  
GCGTGCGCTGACATTTCGCATTTCGCGAGGTTCGGCATTGGCGCAAACCAGCAGCGCTGCAC  
GATCCGTGCCGCCAAGCTCGAGACCGGCAACCCGCGAGCGCCAGATCGCCAAGGACGACG  
TGCTGCTCGTTTCTGGCGGCGCTCGCGGCATCACGCCTCTTTGCATCCGGGAGATCACG  
CGCCAGATCGCGGGCGGCAAGTACATTCTGCTTGGCCGCGAGCAAGGTCTCTGCGAGCGA  
ACCGGCATGGTGCGCTGGCATCACTGACGAGAAGGCTGTGCAAAAGGCTGCTACCCAGG  
AGCTCAAGCGCGCCTTTAGCGCTGGCGAGGGCCCCAAGCCCACGCCCCGCGCTGTCACT  
AAGCTTGTGGGCTCTGTTCTTGGCGCTCGCGAGGTGCGCAGCTCTATTGCTGCGATTGA  
AGCGCTCGGCGGCAAGGCCATCTACTCGTCTGCGACGTGAACTCTGCCGCCGACGTGG  
CCAAGGCCGTGCGCGATGCCGAGTCCCAGCTCGGTGCCCGCGTCTCGGGCATCGTTCAT  
GCCTCGGGCGTGCTCCGCGACCGTCTCATCGAGAAGAAGCTCCCCGACGAGTTTCAGCGC  
CGTCTTTGGCACCAAGGTACCCGGTCTCGAGAACCTCCTCGCCGCCGTGACCCGCGCCA  
ACCTCAAGCACATGGTCCTCTTCAGCTCGCTCGCCGGCTTCCACGGCAACGTCCGGCCAG  
TCTGACTACGCCATGGCCAACGAGGCCCTTAACAAGATGGGCCTCGAGCTCGCCAAGGA  
CGTCTCGGTCAAGTCGATCTGCTTCGGTCCCTGGGACGGTGGCATGGTGACGCCCGCAGC  
TCAAGAAGCAGTTCCAGGAGATGGGCGTGAGATCATCCCCCGCGAGGGCGGCGCTGAT  
ACCGTGGCGCGCATCGTGCTCGGCTCCTCGCCGGCTGAGATCCTTGTCGGCAACTGGCG  
CACCCCGTCCAAGAAGGTTCGGCTCGGACACCATCACCTGCACCGCAAGATTTCCGCCA  
AGTCCAACCCCTTCTCGAGGACCACGTATCCAGGGCCGCCGCGTGCTGCCCCATGACG  
CTGGCCATTGGCTCGCTCGCGGAGACCTGCCTCGGCCTCTTCCCCGGCTACTCGCTCTG  
GGCCATTGACGACGCCCGAGCTCTTCAAGGGTGTCACTGTGACGGCGACGTCAACTGCG  
AGGTGACCCTCACCCCGTCGACGGCGCCCTCGGGCCGCGTCAACGTCCAGGCCACGCTC  
AAGACCTTTTCCAGCGGCAAGCTGGTCCC GGCTACCGCGCCGTATCGTGCTCTCCAA  
CCAGGGCGCGCCCCCGGCCAACGCCACCATGCAGCCGCCCTCGCTCGATGCCGATCCGG  
CGCTCCAGGGCTCCGTCTACGACGGCAAGACCCTCTTCCACGGCCCCGGCCTTCCGCGGC  
ATCGATGACGTGCTCTCGTGACCAAGAGCCAGCTTGTTGGCCAAGTGCAGCGCTGTCCC  
CGGCTCCGACGCCGCTCGCGGCGAGTTTGGCACGGACACTGACGCCCATGACCCCTTCG  
TGAACGACCTGGCCTTTTCAAGCCATGCTCGTCTGGGTGCGCCGACGCTCGGGCCAGGCT  
GCGCTCCCCAACTCGATCCAGCGCATCGTCCAGCACCGCCCGGTCCCGCAGGACAAGCC  
CTTCTACATTACCCTCCGCTCCAACCAGTCGGGCGGTCACTCCAGCACAAAGCACGCC  
TTCAGTTCCACAACGAGCAGGGCGATCTCTTCATTGATGTCCAGGCTTCGGTCATCGCC  
ACGGACAGCCTTGCCTTCTAA

Figure 27 A-2

126/134

TGCCGTCTTTGAGGAGCATGACCCCTCCAACGCCGCCTGCACGGGCCACGACTCCATTT  
CTGCGCTCTCGGCCCGCTGCGGCGGTGAAAGCAACATGCGCATCGCCATCACTGGTATG  
GACGCCACCTTTGGCGCTCTCAAGGGACTCGACGCCTTCGAGCGCGCCATTTACACCGG  
CGCTCACGGTGCCATCCCACTCCCAGAAAAGCGCTGGCGCTTTCTCGGCAAGGACAAGG  
GAGGTGCACTTCCAGCGCCTCCGCACGCCCATGACCCCTGAAGACATGCTCCTCCCTCA  
GCAGCTTCTGGCCGTCACCACCATTGACCGCGCCATCCTCGACTCGGGAATGAAAAAGG  
GTGGCAATGTCGCCGTCTTTGTGCGCCTCGGCACCGACCTCGAGCTCTACCGTCA'CGT  
GCTCGCGTCGCTCTCAAGGAGCGCGTCCGCCCTGAAGCCTCCAAGAAGCTCAATGACAT  
GATGCAGTACATTAACGACTGCGGCACATCCACATCGTACACCTCGTACATTGGGAACC  
TCGTCGCCACGCGCGTCTCGTCGCAGTGGGGCTTACGGGCCCCCTCCTTTACGATCACC  
GAGGGCAACAACTCCGTCTACCGCTGCGCCGAGCTCGGCAAGTACCTCCTCGAGACCGG  
CGAGGTGATGGCGTCTGTTGCGGGTGTGATCTCTGCGGCAGTGCCGAAAACCTTT  
ACGTCAAGTCTCGCCGCTTCAAGGTGTCCACCTCCGATACCCCGCGCGCCAGCTTTGAC  
GCCGCCGCCGATGGCTACTTTGTGCGCGAGGGCTGCGGTGCCTTTGTGCTCAAGCGTGA  
GACTAGCTGCACCAAGGACGACCGTATCTACGCTTGCATGGATGCCATCGTCCCTGGCA  
ACGTCCCTAGCGCCTGCTTGC CGGAGGCCCTCGACCAGGCGCGCGTCAAGCCGGGCGAT  
ATCGAGATGCTCGAGCTCAGCGCCGACTCCGCCCGCCACCTCAAGGACCCGTCCGTCT  
GCCCAAGGAGCTCACTGCCGAGGAGGAAATCGGCGGCCTTCAGACGATCCTTCGTGACG  
ATGACAAGCTCCCGCGCAACGTGCAACGGGCAGTGTCAAGGCCACCGTCGGTGACACC  
GGTTATGCCTCTGGTGCTGCCAGCCTCATCAAGGTGCGCTTTGCATCTACAACCGCTA  
CCTG'GCCAGCAACGGCGACGACTGGGATGAACCCGCCCTGAGGCGCCCTGGGACAGCA  
CCCTCTTTGCGTGCCAGACCTCGCGCGCTTGGCTCAAGAACCCTGGCGAGCGTCGCTAT  
GCGGCCGTCTCGGGCGTCTCCGAGACGCGCTCGTGCTATTCCGTGCTCCTCTCCGAAGC  
CGAGGGCCACTACGAGCGCGAGAACC GCATCTCGCTCGACGAGGAGGCGCCCAAGCTCA  
TTGTGCTTCGCGCCGACTCCCACGAGGAGATCCTTGGTTCGCTCGACAAGATCCGCGAG  
CGCTTCTTGACGCCACGGGCGCCGCCCGCGAGTCCGAGCTCAAGGCGCAGGCCCG  
CCGCATCTTCTCGAGCTCCTCGCGGAGACCCTTGCCCAGGATGCCGCTTCTTCAGGCT  
CGCAAAAGCCCTCGCTCTCAGCCTCGTCTCCACGCCCTCCAAGCTCCAGCGCGAGGTC  
GAGCTCGCGGCCAAGGGTATCCCGCGCTGCTCAAGATGCGCCGCGATTGGAGCTCCCC  
TGCTGGCAGCCGCTACGCGCCTGAGCCGCTCGCCAGCGACCGCGTCGCTTCATGTACG  
GCCAAGGTTCGAGCCCTTACTACGGCATCACCCAAGACATTACCGCATTTGGCCCCGAA  
CTCCACGAGGTCATCAACGAAAAGACGAACCGTCTCTGGGCCGAAGGCGACCGCTGGGT  
CATGCCGCGCGCCAGCTTCAAGTCGGAGCTCGAGAGCCAGCAGCAAGAGTTTGATCGCA  
ACATGATTGAAATGTTCCGTCTTGGATCCTCACCTCAATTGCCTTACCAATCTGGCG  
CGCGACGTTCTCAACATCACGCCCAAGGCCGCTTTGGCCTCAGTCTTGGCGAGATTT  
CATGATTTTTCCTTTTCCAAGAAGAAGGTCTCATCTCCGACCAGCTCACCAGGATC  
TTCGCGAGTCCGACGTGTGGAACAAGGCTCTGGCCGTTGAATTTAATGCGCTGCGCGAG  
GCCTGGGGCATTCCACAGAGTGTCCCCAAGGACGAGTTCTGGCAAGGCTACATTGTGCG  
CGGCACCAAGCAGGATATCGAGGCGGCCATCGCCCCGACAGCAAGTACGTGCGCCTCA  
CCATCATCAATGATGCCAACACCGCCCTCATAGCGGCAAGCCCGACGCCTGCAAGGCT  
GCGATCGCGCGTCTCGGTGGCAACATTCTGCGCTTCCCGTGACCCAGGGCATGTGCGG  
CCACTGCCCCGAGGTGGGACCTTATACCAAGGATATCGCCAAGATCCATGCCAACCTTG

Figure 27 B-1

127/134

AGTTCCCCGTTGTGCGACGGCCTTGACCTCTGGACCACAATCAACCAGAAGCGCCTCGTG  
CCACGCGCCACGGGCGCCAAGGACGAATGGGCCCTTCTTCCTTTGGCGAGTACGCCGG  
CCAGCTCTACGAGAAGCAGGCTAACTTCCCCCAAATCGTCGAGACCATTATACAAGCAAA  
ACTACGACGTCTTTGTGCGAGGTTGGGCCCAACAACCACCGTAGCACCGCAGTGCACACC  
ACGCTTGGTCCCCAGCGCAACCACCTTGCTGGCGCCATCGACAAGCAGAACGAGGATGC  
TTGGACGACCATCGTCAAGCTTGTGGCTTCGCTCAAGGCCACCTTGTTCCCTGGCGTCA  
CGATCTCGCCGCTGTACCACTCCAAGCTTGTGGCGGAGGCTCAGGCTTGCTACGCTGCG  
CTCTGCAAGGGTGAAAAGCCCCAAGAAGAACAAGTTTGTGCGCAAGATTAGCTCAACGG  
TCGCTTCAACAGCAAGGCGGACCCCATCTCCTCGGCCGATCTTGCCAGCTTTCCGCCTG  
CGGACCCTGCCATTGAAGCCGCCATCTCGAGCCGCATCATGAAGCCTGTCGCTCCCAAG  
TTCTACGCGCGTCTCAACATTGACGAGCAGGACGAGACCCGAGATCCGATCCTCAACAA  
GGACAACGCGCCGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTC  
CGTCGCCTGCTCCTTCGGCCCCCGTGCAAAAGAAGGCTGCTCCCGCCGCGGAGACCAAG  
GCTGTTGCTTCGGCTGACGCACTTCGCACTGCCCTGCTCGATCTCGACAGTATGCTTGC  
GCTGAGCTCTGCCAGTGCCCTCCGGCAACCTTGTTGAGACTGCGCCTAGCGACGCCTCGG  
TCATTGTGCCGCCCTGCAACATTGCGGATCTCGGCAGCCGCGCCTTCATGAAAACGTAC  
GGTGTTCGGCGCCTCTGTACACGGGCGCCATGGCCAAGGGCATTGCCTCTGCGGACCT  
CGTCATTGCCGCCGGCCGCGCAGGGCATCCTTGCGTCTTTGGCGCCGGCGGACTTCCCA  
TGCAGGTTGTGCGTGAGTCCATCGAAAAGATTAGGCCGCCCTGCCCAATGGCCCGTAC  
GCTGTCAACCTTATCCATTCTCCCTTTGACAGCAACCTCGAAAAGGGCAATGTGATCT  
CTTCCTCGAGAAGGGTGTACCTTTGTGCGAGGCCTCGGCCTTTATGACGCTCACCCCGC  
AGGTGCTGCGGTACCGCGCGGCTGGCCTCACGCGCAACGCCGACGGCTCGGTCAACATC  
CGCAACCGTATCATTGGCAAGGTCTCGCGCACCGAGCTCGCCGAGATGTTATGCGTCC  
TGCGGCCGAGCACCTTCTTCAGAAGCTCATTGCTTCCGGCGAGATCAACCAGGAGCAGG  
CCGAGCTCGCCCGCCGTGTTCCCGTCGCTGACGACATCGCGGTGCAAGCTGACTCGGGT  
GGCCACACCGACAACCGCCCCATCCACGTCAATTCTGCCCCCTCATCATCAACCTTCGCGA  
CCGCCTTCACCGCGAGTGCGGCTACCCGGCCAACCTTCGCGTCCGTGTGGGCGCCGGCG  
GTGGCATTGGGTGCCCCCAGGCGGCGCTGGCCACCTTCAACATGGGTGCTCCTTTATT  
GTCACCGGCACCGTGAACCAGGTGCGCAAGCAGTCCGGGCACGTGCGACAATGTGCGCAA  
GCAGCTCGCGAAGGCCACTTACTCGGACGTATGCATGGCCCCGGCTGCCGACATGTTG  
AGGAAGGCGTCAAGCTTCAGGTCTCAAGAAGGGAACCATGTTTCCCTCGCGCGCCAAC  
AAGCTCTACGAGCTCTTTTGCAAGTACGACTCGTTGAGTCCATGCCCCCGCAGAGCT  
TGCGCGCGTTCGAGAAGCGCATCTTCAGCCGCGCGCTCGAAGAGGTCTGGGACGAGACCA  
AAAACCTTTTACATTAACCGTCTTCACAACCCGAGAAGATCCAGCGCGCCGAGCGCGAC  
CCCAAGCTCAAGATGTGCTGTGCTTTCGCTGGTACCTGAGCCTGGCGAGCCGCTGGGC  
CAACACTGGAGCTTCCGATCGCGTCATGGACTACCAGGTCTGGTGCGGTCTTGCCATTG  
GTTCCCTTCAACGATTTTCATCAAGGGAACCTTACCTTGATCCGGCCGTGCAACGAGTAC  
CCGTGCGTCTTCAGATTAACAAGCAGATCCTTCGTGGAGCGTGCTTCTTGCGCCGTCT  
CGAAATTCTGCGCAACGCACGCCTTTCGATGGCGCTGCCGCTCTTGTTGGCCAGCATCG  
ATGACACATACGTCCCGGCCGAGAAGCTGTAAGTAAGCTCTCATATATGTTAGTTGCGT  
GAGACCGACACGAAGATAATATCACATACGCTTTTGTGTTGTTCTTTCAATTATTTGTCT  
GTGCTTCATGTTGCTCCTCAGTATCTAGCTGGCGGCTCTTATCTTCTTTTAAATATCT  
GGACAAGGACAAAAACAAGAATAAAGGCGAGAAGATGTGAATTTCAATTCGACTTGAGA

Figure 27 B-2



128/134

ACTCGAAGAGCATTGATGCGGTTAGTATATGGGTATTTTCCAGACACTTTTCATCATCA  
TCATCATCATCATCATTATGAAGAAGTAGTAGCTGATAAAGTAGACTCACTGTTTGCAG  
CGAGAAAAAAAAAAAAAAAAAAAAA

Figure 27 B-3

[illegible]

Figure 27 C-1

130/134

CCAGCTCAACCGCCGCACGGACCAGGGCCAGTACCTCGACGCCGTCGACATTGTCTCCG  
GCAGCGGCAAGAAGAGCCTCGGCTACGCCCACGGTCCAAGACGGTCAACCCGAACGAC  
TGGTTCTTCTCGTGCCACTTTTGGTTTGACTCGGTCATGCCCAGGAAGTCTCGGTGTCTGA  
GTCCATGTTCCAGCTCGTCGAGGCCATCGCCGCCACGAGGATCTCGCTGGCAAAGCAC  
GGCATTGCCAACCCACCTTTGTGCACGCCCCCGGCAAGATCAAGCTGGAAGTACCGC  
GGSCAGCTCACGCCCAAGAGCAAGAAGATGGACTCGGAGGTCCACATCGTGTCCGTGGA  
CGCCACGACGGCGTTGTGACCTCGTCGCCGACGGCTTCCTCTGGGCCGACAGCCTCC  
GCGTCTACTCGGTGAGCAACATTGCGGTGCGCATCGCCTCCGGTGAGGCCCCCTGCCGCC  
GCCTCCTCCGCCGCTCTGTGGGCTCCTCGGCTTCGTCCGTGAGCGCACGCGCTCGAG  
CCCCGCTGTGCGCTCCGGCCCCGGCCAGACCATCGACCTCAAGCAGCTCAAGACCGAGC  
TCCTCGAGCTCGATGCCCGCTCTACCTCTCGCAGGACCCGACCAGCGGCCAGCTCAAG  
AAGCACACCGACGTGGCCTCCGGCCAGGCCACCATCGTGACGCCCTGCACGCTCGGCGA  
CCTCGGTGACCGCTCCTTCATGGAGACCTACGGCGTCGTGCCCCGCTGTACACGGGCG  
CCATGGCCAAGGGCATTGCCTCGGCGGACCTCGTCATCGCCGCCGGCAAGCGCAAGATC  
CTCGGCTCCTTTGGCGCCGGCGGCCTCCCCATGCACCACGTGCGCGCCGCCCTCGAGAA  
GATCCAGGCCGCCCTGCCTCAGGGCCCCCTACGCCGTCAACCTCATCCACTCGCCTTTTG  
ACAGCAACCTCGAGAAGGGCAACGTGATCTCTTCCTCGAGAAGGGCGTCACTGTGGTG  
GAGGCCTCGGCATTATGACCCTACCCCGCAGGTGTCGCGCTACCGCGCCGCCGCGCT  
CTCGCGCAACGCCGACGGTTCGGTCAACATCCGCAACCGCATCATCGGCAAGGTCTCGC  
GCACCGAGCTCGCCGAGATGTTTCATCCGCCCGGCCCCGGAGCACCTCCTCGAGAAGCTC  
ATCGCCTCGGGCGAGATCACCCAGGAGCAGGCCGAGCTCGCGCGCCGCGTTCCTCGTCGC  
CGACGATATCGCTGTGAGGCTGACTCGGGCGGCCACACCGACAACCGCCCCATCCACG  
TCATCCTCCCGCTCATCATCAACCTCCGCAACCGCCTGCACCGCGAGTGCGGCTACCCC  
GCGCACCTCCGCGTCCGCGTTGGCGCCGGCGGTGGCGTCCGCTGCCCGCAGGCCGCCGC  
CGCCGCGCTCACCATGGGCGCCGCTTCATCGTCAACCGGCACTGTCAACCAGGTGCGCA  
AGCAGTCCGGCACCTGCGACAACGTGCGCAAGCAGCTCTCGCAGGCCACCTACTCGGAT  
ATCTGCATGGCCCCGGCCGCCGACATGTTTCGAGGAGGGCGTCAAGCTCCAGGTCTCAA  
GAAGGGAACCATGTTCCCTCGCGGCCAACAAGCTCTACGAGCTCTTTTGCAAGTACG  
ACTCCTTCGACTCCATGCCTCCTGCGGAGCTCGAGCGCATCGAGAAGCGTATCTTCAAG  
CGCGCACTCCAGGAGGTCTGGGAGGAGACCAAGGACTTTTACATTAACGGTCTCAAGAA  
CCCGGAGAAGATCCAGCGCGCCGAGCACGACCCCAAGCTCAAGATGTCGCTCTGCTTCC  
GCTGGTACCTTGGTCTTGCCAGCCGCTGGGCCAACATGGGCGCCCCGGACCGCGTCATG  
GACTACCAGGTCTGGTGTGGCCCGGCCATTGGCGCCTTCAACGACTTCATCAAGGGCAC  
CTACCTCGACCCCGCTGTCTCCAACGAGTACCCCTGTGTGTCGTCAGATCAACCTGCAAA  
TCCTCCGTGGTGCCTGCTACCTGCGCCGTCTCAACGCCCTGCGCAACGACCCGCGCATT  
GACCTCGAGACCGAGGATGCTGCCTTTGTCTACGAGCCACCAACGCGCTCTAAGAAAG  
TGAACCTTGTCCTAACCCGACAGCGAATGGCGGGAGGGGGCGGGCTAAAAGATCGTATT  
ACATAGTATTTTTCCCCTACTCTTTGTGAAAAAAAAAAAAAAAAAAAAA

Figure 27 C-2

131/134

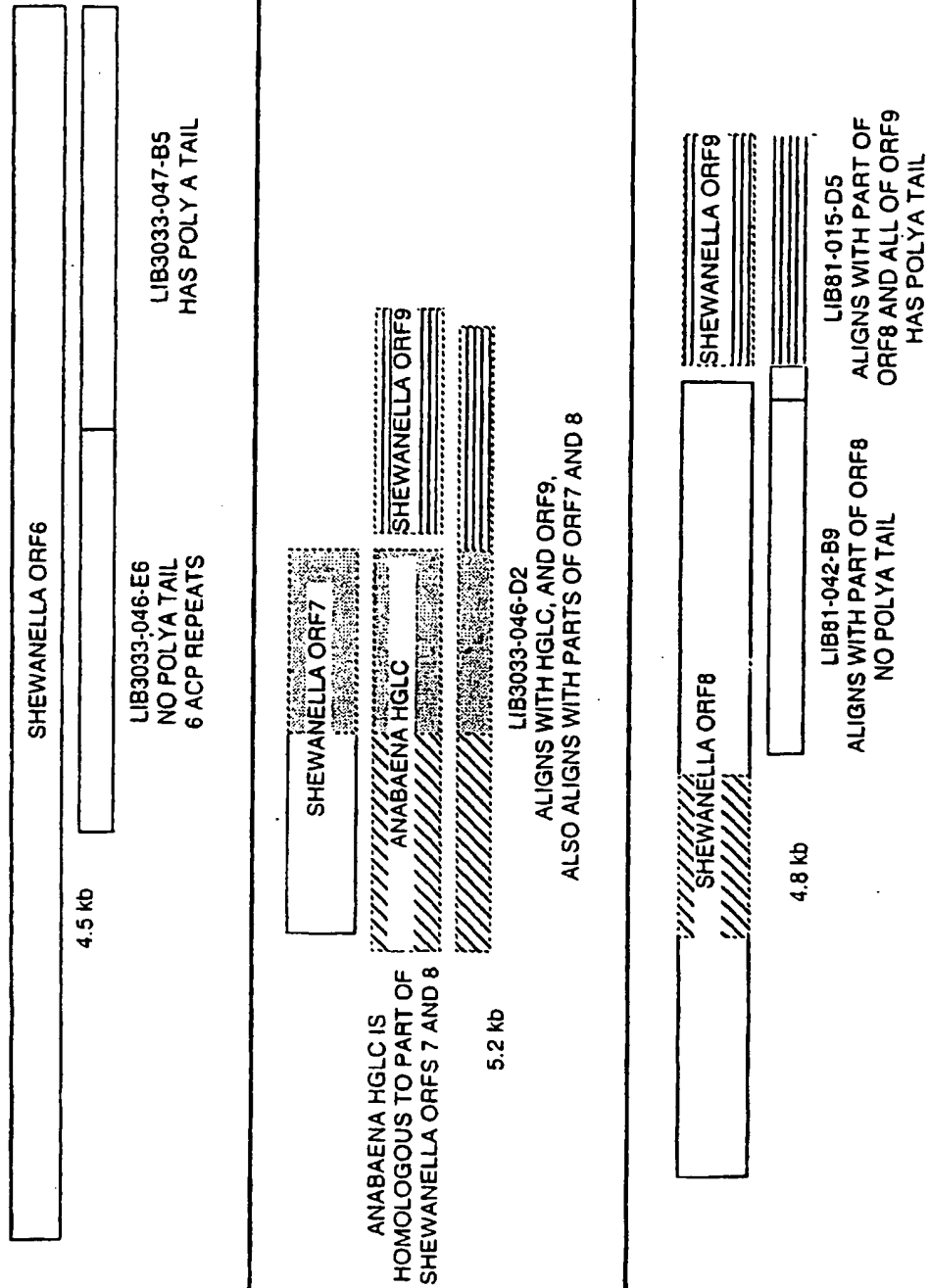


Figure 28

132/134

RCRRVSPRRAAPPPPLARTPARLAAPAVSNELLEKAETVVMEVLAAKTGYETDMIESDM  
ELETELGIDSIKRVEILSEVQAMLNVEAKDVDALSRTTRTVGEVVNAMKAEIAGGSAPAP  
AAAAPGPAAAAPAPAVSSELLEKAETVVMEVLAAKTGYETDMIESDMELETELGIDSIK  
RVEILSEVQAMLNVEAKDVDALSRTTRTVGEVVDAMKAEIAGSSASAPAAAAPAPAAAAP  
APAAAAPAVSNELLEKAETVVMEVLAAKTGYETDMIESDMELETELGIDSIKRVEILSE  
VQAMLNVEAKDVDALSRTTRTVGEVVDAMKAEIAGGSAPAPAAAAPAPAAAAPAVSNELL  
EKAETVVMEVLAAKTGYETDMIESDMELETELGIDSIKRVEILSEVQAMLNVEAKDVDA  
LSRTTRTVGEVVDAMKAEIAGSSAPAPAAAAPAPAAAAPAPAAAAPAVSSELLEKAETVV  
MEVLAAKTGYETDMIESDMELETELGIDSIKRVEILSEVQAMLNVEAKDVDALSRTTRTV  
GEVVDAMKAEIAGGSAPAPAAAAPAPAAAAPAVSNELLEKAETVVMEVLAAKTGYETDM  
IESDMELETELGIDSIKRVEILSEVQAMLNVEAKDVDALSRTTRTVGEVVDAMKAEIAGG  
SAPAPAAAAPASAGAAPAVKIDSVHGADCDDLMLHAKVVDIRRPDELILERPENRPVL  
VDDGSELTLALVRVLGACAVVLTFFGLQLAQRAGAAAIRHVLAKDLAESAIEKAIKEA  
EQRFGLGGFISQQAERFEPAEILGFTLMCAKFAKASLCTAVAGGRPAFIGVARLDGRL  
GFTSQGTSALKRAQRGAIFGLCKTIGLEWSESDVFSRGVDIAQGMHPEDAIVREM  
ACADIRIREVGIGANQQRCTIRAAKLETGNPQRQIAKDDVLLVSGGARGITPLCIREIT  
RQIAGGKYILLGRSKVSASEPAWCAGITDEKAVQKAATQELKRAFSAGEGPKPTPRAVT  
KLVGSLVGLAREVRSSIAAIEALGGKAIYSSCDVNSAADVAKAVRDAESQLGARVSGIVH  
ASGVLRDRLEIEKKLPDEFDAVFGTKVTGLENLLAAVDRANLKHMLVFSSLAGFHGNVQ  
SDYAMANEALNKMGLELAKDVS VKSICFGPWDGGMVTPQLKKQFQEMGVQIIPREGGAD  
TVARIVLGSSPAEILVGNWRTPSKKVGSDTITLHRKISAKSNPFLEDHVIQGRRLPMT  
LAIGSLAETCLGLFPGYSLWAIDDAQLFKGVTVDGVDNCEVTLPSTAPSGRVNVQATL  
KTFSSGKLVPAYRAVIVLSNQGAPPANATMQPPSLDADPALQGSVYDGKTLFHGPAFRG  
IDDVLSCTKSQLVAKCSAVPGSDAARGEATDTDAHDPFVNDLAFQAMLVWVRRTLGOA  
ALPNSIQRIVQHRPVPQDKPFYITLRSNQS GGHSQHKHALQFHNEQGDLFIDVQASVIA  
TDSLAF

Figure 29 A

133/134

AVFEEHDP SNAACTGHDSISALSARCGGESNMRIAITGMDATFGALKGLDAFERAIYTG  
AHGAIPLEKRWRF LGKDKDFLDLCGVKATPHGCYIEDVEVDFQRLRTPMTPEDMLLPQ  
QLLAVTTIDRAILD SGMKKGNVAVFVGLGTDLELYRHRARVALKERVRPEASKKLNDM  
MQYINDCGTSTSYTSYIGNLVATRVSSQWGFTGPSFTITEGNNSVYRCAELGKYLLETG  
EVDGVVVAGVDLCGSAENLYVKSRRFKVSTSDTPRASFDAAADGYFVGEGCGAFVLKRE  
TSCTKDDRIYACMDAIVPGNVPSACLREALDQARVKPGDIEMLELSADSARHLKDPSVL  
PKELTAE E E I GGLQ TILRDDD KLPRNVATG SVKATV GDTGYASGAASLIKAALCIYNRY  
LPSNGDDWDEPAPEAPWDSTL FACQTSRAWLKNPGERRYAAVSGVSETRSCYSVLLSEA  
EGHYERENRISLDEEAPKLIVLRADSHEEILGRLDKIRERFLQPTGAAPRESELKAQAR  
RIFLELLGETLAQDAASSGSQKPLALSLVSTPSKLQREVELAAKGIPRCLKMRRDWSSP  
AGSRYAPEPLASDRVAFMYGEGRSPYYGITQDIHRIWPHELHEVINEKTNRLWAECDRWV  
MPRASFKSELESQQQEFDRNMIEMFRLGILTSIAFTNLARDVLNITPKAAFGLSLGEIS  
MIFAFSKKNGLISDQLTKDLRESDVWNKALAVEFNALREAWGIPQSVPKDEFWQGYIVR  
GKQDIEAAIAPDSKYVRLTIINDANTALISGKPDACKAAIARLGGNIPALPVTQGMCG  
HCPEVGPyTKDIAKIHANLEFPVVDGLDLWTTINQKRLVPRATGAKDEWAPSSFGEYAG  
QLYEKQANFPQIVETIYKQNYDV FVEVGPNNHRSTAVRTTLGPQRNHLA GAIDKQNE DA  
WTTIVKLVASLKAHLVPGVTISPLYHSLVAEAQACYAALCKGEKPKKNKFVRKIQLNG  
RFNSKADPISSADLASFPADPAIEAAISSRIMKPVAPKFYARLNIDEQDETRDPI LNK  
DNAPSSSSSSSSSSSSSSSSPSPAPSAPVQKKAAPAAETKAVASADALRSALLDLDSMLA  
LSSASASGNLVETAPSDASVIVPPCN IADLGSRAFMKTYGVSAPLYTGAMAKGIASADL  
VIAAGRQGILASFGAGGLPMQVVRESIEKIQAALPNGPYAVNLIHSPFDSNLEKGNVDL  
FLEKGVTFVEASAFMTLTPQVVRYRAAGLTRNADGSVNIRNRIIGKVSRTELAEMFMRP  
APEHLLQKLIASGEINQEQAELARRVPVADDIAVEADSGGHTDNRPIHVILPLIINLRD  
RLHRECGYPANLRVRVGAGGGIGCPQAALATFNMGASFIVTGTVNQVAKQSGTCDNVRK  
QLAKATYSDVCMAPAADMFEEGVKLQVLKKGTMFPSRANKLYELFCKYDSFESMPPAEL  
ARVEKRIFSRALEEVWDETKNFYINRLHNPEKIQRAERDPKLMKSLCFRWYLSLASRWA  
NTGASDRVMDYQVWCCGPAIGSFNDFIKGTYLDPAVANEYPCVVQINKQILRGACFLRRL  
EILRNARLSDGAAALVASIDDTYVPAEKL

Figure 29 B

134/134

RAEAGREPEPAPQITSTAAESQQQQQQQQQQQQQQQPPREGDKEKAAETMALRVKTNKKPCWEMT  
KEELTSGKTEVFNYEELLEFAEGDIAKVFGEFAVIDKYPRRVRLPAREYLLVTRVTLMDAEVN  
NYRVGARMVTEYDLPVNGELSEGDCPWAVLVESGQCDLMLISYMGIDFQNOGDRVYRLNNTTL  
TFYGVAHEGETLEYDIRVTGFAKRLDGGISMFFFYDCYVNGRLLIEMRDGCAGFFTNEELDAG  
KGVVFTRGDLAARAKIPKQDVSPYAVAPCLHKTCLNEKEMQTLVDKDWASVFGSKNGMPEINYK  
LCARKMLMIDRVTSIDHKGGVYGLGQLVGEKILERDHWYFPCHFVKDQVMAGSLVSDGCSQMLK  
MYMIWLGLHLTTGPFDFRPVNGHPNKNVRCRQISPHKGLVYVMEIKEMGFDEDNDPYAIADVNI  
IDVDFEKGQDFSLDRISDYGKGDNLNKKIVDFKGIALKMQRSTNKNPSKVQPVFANGAATVG  
PEASKASSGASASASAAPAKPAFSADVLAPKPVALPEHILKGDALAPKEMSWHPMARIPGNPTP  
SFAPSAYKPRNIAFTPPFGNPNDNDHTPGKMPLTWFNMAEFMAGKVSMCLGPEFAKPDSDNTSR  
SPAWDLALVTRAVSVSDLKHVNRYRNIDLPSKGMTVGEFDCPADAWFYKGACNDAHMPYSILME  
IALQTSGVLTSVLKAPLTMKDDILFRNLDANAEFVRADLDYRGKTI RNVTKCTGYSMLGEMGV  
HRFTFELYVDDVLFYKGSTSFGEVFPVFAAQAGLDNGRKSEPFIEKNVPASQVSSFDVRPNG  
SGRTAIFANAPSGAQLNRRTDQGQYLDIVSGSGKSLGYAHGSKTVNPNDWFFSCHFWFDS  
VMPGSLGVESMFQLVEAIAAHEDLAGKARHCQPHLCARPRARSSWKYRGQLTPKSKKMDSEVHI  
VSVDADHDGVVDLVADGFLWADSLRVYSVSNIRVRIASGEAPAAASSAASVGSSASSVERTRSSP  
AVASGPAQTIDLKQLKTELLELDAPLYLSQDPTSGQLKKHTDVASGQATIVQPCTLGDLGDRSF  
METYGVVAPPLYTGAMAKGIASADLVIAAGKRKILGSFGAGGLPMHHVRAALEKIQAAALPQGPYA  
VNLHSPFDSNLEKGNVDLFLEKGVTVVEASAFMTLTTPQVVRYRAAGLSRNADGSVNIRNRIIG  
KVSRTELAEMFIRPAPEHLLLEKLIASGEITQEQAELARRVPVADDIAVEADSGGHTDNRPIHVI  
LPLIINLRNRLHRECGYPAPHLRVRVGAGGGVGCPOAAAAALTMGAAFIVTGTVNQVAKQSGTCD  
NVRKQLSQATYSIDICMAPAADMFEEGVKLQVLKKGTMFPSRANKLYELFCKYDSFDSMPPAELE  
RIEKRIFKRALQEVWEETKDFYINGLKNPEKIQRAEHPKCLKMSLCFRWYLG LASRWANMGAPD  
RVMDYQVWCGPAIGAFNDFIKGTYLDPAVSNEYPCVVQINLQILRGACYLRRRLNALRNDPRIDL  
ETEDAAFFVYEPTNAL

Figure 29 C

## SEQUENCE LISTING

<110> Lassner, Michael  
Metz, James G  
Facciotti, Daniel

<120> SCHIZOCHYTRIUM PKS GENES  
<130> CGNE.131.02WO

<140> Not Yet Assigned  
<141> 2000-01-14

<150> 09/231,899  
<151> 1999-01-14

<150> 09/090,793  
<151> 1998-06-04

<150> 60/048,650  
<151> 1997-06-04

<160> 86  
<170> PatentIn Ver. 2.0

<210> 1  
<211> 37895  
<212> DNA  
<213> *Shewanella putrefaciens*

<400> 1  
gatctcttac aaagaaacta tctcaatgtg aatttaacct taattccggt taattacggc 60  
ctgatagagc atcaccacaat cagccataaa actgtaaagt ggggtactcaa aggtggctgg 120  
gcgattcttc tcaaatacaa agtgcccaac ccaagcaaat ccataccga taacaggtaa 180  
aagtagcaat aaaccccagc gctgagttag taatacataa gcgaataata ggatcactaa 240  
actactgccg aaatagtgtg atattcgaca gtttctatgc tgatgttgag ataaataaaa 300  
agggtaaaat tcagcaaaaag aacgatagcg cttactcatt actcacacct cggtaaaaaa 360  
gcaactcgcc attaaacttg ccaatcgta gttgttctat cgtctcaaag ttatgccgac 420  
taaataactc tatatgtgca ttatgattag caaaaactcc gataccatca agatgaagtt 480  
gttcacacac ccaactcaaa actgcgtcga taagcttact gccatagccc ttgccttgct 540  
ccacatttgc gatagcaata aactgtaaaa tgccacattg gccacttggt aagctctcta 600  
taatctgatt ttctttgtta ataagtgcct gagttgaata ccaaccagta cttacaaca 660  
tctttaaacg ccaatgccaa aaacgcgctt caccataagg aacctgctga gtcactatgc 720  
aggctacgcc tatcaatcta tccccaacga acataccaat aagtgcctgc tctgttgcc 780  
agagctcatt gagttcttct cgaatagccc cgcgaagctt ttgctcatcc tgcgcttgat 840  
caccactaaa aagtgtttcg ataaaaaagg gatcatcatg ataggcggtg tagagaatag 900  
aggctgctat gcgtaaatct tctgccgtga gataaactgc acgacactct tccatggctt 960  
gatcttccat tgttattgtc cttgaccttg atcacacaac accaatgtaa caagactgta 1020  
tagaagtgca attaataatc aattcgtgca ttaagcaggt cagcatttct ttgctaaaca 1080  
agcttctattg gctttgacaa aactttgcct agactttaac gatagaaatc ataataaag 1140



agaaaagcta caacctagag gggaataatc aaacaactgc taagatctag ataatgtaat 1200  
 aaacaccgag ttatcgcacc atacttagat agagtcatag caacgagaat agttatggat 1260  
 acaacgccgc aagatctatc acacctgttt ttacagctag gattagcaaa tgatcaaccc 1320  
 gcaattgaac agtttatcaa tgaccatcaa ttagcggaca atatattgct acatcaagca 1380  
 agcttttggg gcccatcgca aaagcacttc ttaattgagt catttaatga agatgccag 1440  
 tggaccgaag tcacgcacca cttagacacc ttattaagaa aaaactaacc attacaacag 1500  
 caactttaaa ttttgccgta agccatctcc cccacccca caacagcgtt gttgcttatg 1560  
 accactggag tacattcgtc tttagtcgtt ttaccatcac catgggtacg ttgagtgcga 1620  
 taaaaaagca cataaacttc tttatcggcc tgaatatagg cttcgttaaa atcagctgtt 1680  
 cccattaaag taaccacttg ctctttactc atgcctagag atatctttgt caaattgtca 1740  
 cggtttttat cttgagtttt ctcccaagca ccgtgattat ccagtcaga ttcccatca 1800  
 ccaacattga ccacacagcc cgttagccct aagcttgcaa tcccaaaaca tgctaaacct 1860  
 aataatttat ttttcatttt aacttcctgt tatgacatta tttttgctta gaagaaaagc 1920  
 aacttacatg ccaaaacaca agctgttgtt ttaaatgact ttatttatta ttagcctttt 1980  
 aggatagcc tagagcaata ataattacca atgtttaagg aatttgacta actatgagtc 2040  
 cgattgagca agtgctaaca gctgctaaaa aaatcaatga acaaggtaga gaaccaacat 2100  
 tagcattgat taaaaccaa cttggtaata gcacccaat gcgcgagtta atccaaggtt 2160  
 tgcaacagtt taagtctatg agtgcagaag aaagacaagc aatacctagc agcttagcaa 2220  
 cagcaaaaaga aactcaatat ggtcaatcaa gcttatctca atctgaacaa gctgatagga 2280  
 tcctccagct agaaaacgcc ctcaatgaat taagaaacga atttaattggg ctaaaagtc 2340  
 aatttgataa cttaacaaca aacctgatga ataaagagcc tgacaccaa tgcatgtaat 2400  
 tgaactacga tttgaatgtt ttgataacac cagcattact gcagcagaaa aagccattaa 2460  
 tggtttgctt gaagcttatc gagccaatgg ccagggttcta ggctcgtgaat ttgccgttgc 2520  
 atttaacgat ggtgagttta aagcacgcat gttaacccca gaaaaaagca gcttatctaa 2580  
 acgctttaat agtccttggg taaatagtgc actcgaagag ctaaccgaag ccaattgct 2640  
 tgcgccacgt gaaaagtata ttggccaaga tattaattct gaagcatcta gccaaagacac 2700  
 accaagttag cagctacttt acacaagtta tgtgcacatg tgctcaccac taagaaatgg 2760  
 cgacaccttg cagcctattc cactgtatca aattccagca actgccaacg gcgatcataa 2820  
 acgaatgatc cgttggcaaa cagaatggca agcttgtgat gaattgcaa tggccgcagc 2880  
 tactaaagct gaatttgccg cacttgaaga gtaaccagt catcagagtg atctatttag 2940  
 gcgtgggttg gacttacgtg gcagagtcga atacttgacg aaaattccga cctattacta 3000  
 tttataccgt gttgpcggtg aaagcttagc agtagaaaag cagcgtctct gtcctaagt 3060  
 tggcagtcga gaatggctgc tcgataaacc attattggat atgttccatt ttcgctgtga 3120  
 cacctgccgc atcgtatcta atatctcttg ggaccattta taactcttcc gagtcttacc 3180  
 acactagagt ttagtcagca taaaaatggc gcttatattt caattaaaag aaatataagc 3240  
 gccattttca tcgatactat atatcagcag actattttcc gcgtaaatta gccacatta 3300  
 atttcattct ttgccagatc cctggatgat ctagtgtgg catcgactct tcaataggtt 3360  
 taaccgcagg tgtaaccctt ggagtcaatt cgtttataaa ctcgtttaaa ctgtcactta 3420  
 atttaacgct ttgtacttca cctggaattt caatccatac gctgccatca ctattattaa 3480  
 ccgtcaacat tttatcttca tcacaaagaa taccaataaa ccaagtcggc tcttgcttaa 3540  
 gctttctctt catcattaaa tgaccaatga tgttttgttg taagtattca aaatcagttt 3600  
 gatccacac ttggattagc tcaccttggc cccattgtga gtcaaaaaat agcgggtgcag 3660  
 aaaaatgact gccaaaaaat ggattaattt ctgcagataa tgtcatttca agtgcgtgtt 3720  
 caacattagc aaattcacca ggttgttgac gtacaaccga ttgccaaaac actgcgccat 3780  
 cggagccgc ttcggcgaca acacactcag acttttgtcc ttgcgcataa tatcttggct 3840  
 gttcaccaag cttatccatg taggcttgtt gatatttaga taaaaaaga tctaaagcag 3900  
 gtaaagaaga cacttaagcc agttccaaaa tcagttataa taggggtcta ttttgacatg 3960  
 gaaaccgtat tgatgacaca acatcatgat ccctacagta acgccccga actttctgaa 4020

ttaacttttag gaaagtcgac cggttatcaa gagcagtatg atgcatcttt actacaagcg 4080  
 tgccgcgtaa attaaaccgt gatgctatcg gtctaaccaa tgagctacct ttatcatggct 4140  
 gtgatatttg gactggctac gaactgtctt ggctaaatgc taaaggcaag ccaatgattg 4200  
 ctattgcaga ctttaacctt agttttgata gtaaaaatct gatcgagtct aagtcgttta 4260  
 agctgtattt aaacagctat aaccaaaccac gatttgatag cgttcaagcg gttcaagaac 4320  
 gtttaactga agacttaagc gcctgtgccc aaggcacagt tacggtaaaa gtgattgaac 4380  
 ctaagcaatt taaccacctg agagtgggtg atatgccagg tacctgcatt gacgatttag 4440  
 atattgaagt tgatgactat agctttaact ctgactatct caccgacagt gttgatgaca 4500  
 aagtcaggtg tgctgaaacg ctaacgtcaa acttattgaa atcaaactgc ctaatcactt 4560  
 ctcagcctga ctggggtaca gtgatgatcc gttatcaagg gcctaagata gaccgtgaaa 4620  
 agctacttag atatctgatt tcatttagac agcacaatga atttcatgag cagtgtgttg 4680  
 agcgtatatt tgttgattta aagcactatt gccaatgtgc caaacttact gtctatgcac 4740  
 gttatacccg ccgtggtggt ttagatatca acccatatcg tagcgacttt gaaaaccctg 4800  
 cagaaaatca gcgcctagcg agacagtaat tgattgcagt acctacaaaa aacaatgcct 4860  
 ataagccaag cttatgggca tttttatatt atcaacttgt catcaaacct cagccgcaa 4920  
 gccttttagt tttatcgcta aattaagccg ctctctcagc caaatatttg caggattttg 4980  
 ctgtaattta tggctccaca ccatgaaata ctctatcggc tctaccgcaa aaggtaagtc 5040  
 aaatacctgt aagccaaaca gcttggcata ttcgtcagtg tgggcttttg acgcgatagc 5100  
 taacgcatca ctttttgagg caaccgacat catacttaat attgatgatt gctcgctgtg 5160  
 catttgccct gccggtaca cctgtttagt cagcaagtcg gcaacactta aattgtagcg 5220  
 gcgcatctta aaaataatat gcttttcatt aaagtattgc tcttgctca acccaccttg 5280  
 gatccttggg tgagcatttc gtgccacaca aactaattta tcctgcatta ctttttgact 5340  
 cttaaagccc gcagattctg gcagccaaat atctaaggct aaatccacct tttctagttg 5400  
 taggtccatc tgcaactctt cttcaatgag cggcggtcca cgaaatacaa tattaattgc 5460  
 agtgccctgt aacacttgct caatttgatc ttgcaagagt tgtattgccg actcgctggc 5520  
 atacacataa aaagttcgct cacttgaagt ggggtcaaat gcttcaaagc tagtcgcaac 5580  
 ttgctcaatt gttgacatag cgcccgcgag ctgttgataa agcgtcatcg cacttgcggt 5640  
 aggtttaact cccctaccca ctcgagtaaa caactcttct ccaacaatac tttttagcct 5700  
 cgaaatcgca ttactaacg acgactgagt caaatccagc tcttctgccg cccggctaaa 5760  
 agatgaggtg cgatacacg cagtaaaaaa gcgaaataaa ttaagatcaa aagctttttg 5820  
 ctgcgacata aatcagctat ctccctatcc ttatccctat ccttataaaa agttagctcc 5880  
 agagcactct agctcaaaaa caactcagcg tattaagcca atattttggg aactcaatta 5940  
 atattcataa taaaagtatt cataatataa ataccaagtc ataatttagc cctaattatt 6000  
 aatcaattca agttacctat actggcctca attaaagcaa tgtctcatca gtctccctgc 6060  
 aactaaatgc aatattgaga cataaagctt tgaactgatt caatcttacg agggtaactt 6120  
 atgaaacaga ctctaattggc tatctcaatc atgtcgcttt tttcattcaa tgcgctagca 6180  
 gcgcaacatg aacatgacca catcactgtt gattacgaag ggaaagccgc aacagaacac 6240  
 accatagctc acaaccaagc ttagctaaa aacttaact ttgccgacac gcgtgcattt 6300  
 gagcaatcgt ctaaaaatct agtcgccaaag tttgataaag caactgccga tatattacgt 6360  
 gccgaatttg cttttattag cgatgaaatc cctgactcgg ttaaccgcgc tctctaccgt 6420  
 caggctcagc ttaatatggt gcctaattgg ctgtataaag tgagcgatgg catttaccag 6480  
 gtccgcggta ccgacttatc taaccttaca cttatccgca gtgataacgg ttggatagca 6540  
 tacgatgttt tgtaaccaaa agaagcagca aaagcctcac tacaatttgc gttaaagaat 6600  
 ctacctaaag atggcgattt acccgttggt gcgatgattt actcccatag ccatgaggac 6660  
 cactttggcg gagctcgcgg tgttcaagag atgttccctg atgtcaaagt ctacggctca 6720  
 gataacatca ctaaagaaat tgctgatgag aacgtacttg ccggtaacgc catgagccgc 6780  
 cgcgagctt atcaatacgg cgcaacactg ggcaaacatg accacgggtat tgttgatgct 6840  
 gcgctaggta aaggtctatc aaaagtgaa atcacttacg tcgccccaga ctacacctta 6900

aacagtgaag gcaaatggga aacgctgacg attgatggtc tagagatggt gtttatggat 6960  
 gcctcgggca ccgaagctga gtcagaaatg atcacttata ttcctctaa aaaagcgctc 7020  
 tggacggcgg agcttaccta tcaaggatg cacaacattt atacgctgcg cggcgctaaa 7080  
 gtacgtgatg cgctcaagtg gtcaaaagat atcaacgaaa tgatcaatgc ctttggtcaa 7140  
 gatgtcgaag tgctgtttgc ctgcactct gcgccagtgt ggggtaacca agcgatcaac 7200  
 gatttcttac gcctacagcg tgataactac ggcctagtgc acaatcaaac cttgagactt 7260  
 gccaacgatg gtgtcggat acaagatatt ggcgatgcga ttcaagacac gattccagag 7320  
 tctatctaca agacgtggca taccaatggt taccacggca cttatagcca taacgctaaa 7380  
 gcggtttata acaagtatct aggctacttc gatatgaacc cagccaacct taatccgctg 7440  
 ccaaccaagc aagaatctgc caagtgtgtc gaatacatgg gcggcgcgaga tgcgcgaatt 7500  
 aagcgcgcta aagatgatta cgctcaagggt gaataccgct ttgttgcaac ggcattaaat 7560  
 aaggtggtga tggccgagcc agaaaatgac tccgctcgtc aattgctagc cgatacctat 7620  
 gagcaacttg gttatcaagc agaaggggct ggctggagaa acatttactt aactggcgca 7680  
 caagagctac gagtaggtat tcaagctggc gcgcctaaaa ccgcatcggc agatgtcatc 7740  
 agtgaaatgg acatgccgac tctatttgac ttcctcgcgg tgaagattga tagtcaacag 7800  
 gcggctaagc acggcttagt taagatgaat gttatcacc ctgatactaa agatattctc 7860  
 tatattgagc taagcaacgg taacttaagc aacgcagtg tgcacaaaga gcaagcagct 7920  
 gacgcaaacc ttatggttaa taaagctgac gttaaccgca tcttacttgg ccaagtaacc 7980  
 ctaaaagcgt tattagccag cgcgcatgcc aagctcactg gtgataaaac ggcatttagt 8040  
 aaaatagccg atagcatggt cgagtttaca cctgacttcg aaatcgtaac aacgcctgtt 8100  
 aaatgaggca ttaatctcaa caagtgaag ctagacataa aaatggggcg attagacgcc 8160  
 ccatTTTTtT tgcaattttg aactagctag tcttagctga agctcgaaca acagctttaa 8220  
 aattcacttc ttctgctgca atacttattt gctgaactg accaatactc agtgcaaaac 8280  
 gtaactatc atcaagatgg ccagtaaac aatgccaatt atcagcagcg ttcatTTgtt 8340  
 gttctttagc ctcaatcaaa cctaaaccag acttttTgtg ctcagcgTta ggcttattag 8400  
 aactcgactc tagtaaagca agaccaatat cttgttttaa caaaacctgt cgctgattaa 8460  
 gttgatgctc aacctTgtga tccgcaatag catcggaat atcaacacaa tggctcaagc 8520  
 ttttaggtgc attaaTcca agaaaagttt cgctcagTgc agagaagtca aacgcaaaag 8580  
 atttttagcga taatgccagc ccaagtcctt tcgctttaat gtaagactcc ttgagcgccc 8640  
 acaaatcaaa aaagcggTct cgctgcaagg cctctggtaa cgctaacaag gctcgctttt 8700  
 ctgattcaga gaaataatga ctaagaatag agtgatatt ggtgctgtta cggcaacgct 8760  
 caatgtcgac gccaaactca atactagcag agtcagttt ctccttTgct gcctgactgg 8820  
 cgcttttatt atcagcagtg caaatgccta ctaatagcca atctccacta tgactcacat 8880  
 taaagtggac cccggtttga gcaaatTgc catcactcaa tctaggctta ctttTgtgc 8940  
 catattcaaa gcgccattca ttggggcgta tttcactatg ttgtgacaat aaagcgcgca 9000  
 aatagcctct taccattaaa cttgagttt tagcttctt tttaatgtag cgattaaacct 9060  
 taattaactc atcttcaggc agccatgact taaccaactc tgtagtctgg ttatcgact 9120  
 cttgtattgt taacggacag aagtataagg aaatcaatcg agaagttagc aatttttcag 9180  
 gacactcttt aaagcaacaa acataacccc tatttttacc aatttaagat caaaactaaa 9240  
 gccaaaacta attgagaata gtgtcaaact agcttttaaag gaaaaaata taaaaagaa 9300  
 attatacttg tataaattat ttacacacc aaagccatga tcttcacaaa attagctccc 9360  
 tctccctaaa acaagattga ataaaaaat aaaccttaac tttcatatag ataaaaacaa 9420  
 ccaatgggat aaagtatatt gaattcattt ttaaggaaaa attcaaattg aattcaagct 9480  
 cttcagtaaa agcatatttt gccgttagtg tgaaaaaaa caaatttaaa aaccaacata 9540  
 gaacaaataa gcagacaata aaaccaaggc gcaacacaaa caacgcgctt acaattttca 9600  
 caaaaaagca acaagagtaa cgtttagtat ttggatatgg ttattgtaat tgagaatttt 9660  
 ataacaatta tattaaggga atgagtatgt ttttaaattc aaaactttcg cgctcagTca 9720  
 aacttgccat atccgcaggc ttaacagcct cgctagctat gcctgttttt gcagaagaaa 9780

ctgctgctga agaacaata gaaagagtcg cagtgaccgg atcgcgaaatc gctaaagcag 9840  
 agctaactca accagctcca gtcgtcagcc tttcagccga agaactgaca aaatttggtgta 9900  
 atcaagattt aggtagcgta ctagcagaat tacctgctat tgggtgcaacc aacactatta 9960  
 ttggttaataa caatagcaac tcaagcgag gtgttagctc agcagacttg cgtcgtctag 10020  
 gtgctaacag aaccttagta ttagtcaacg gtaagcgcta cgttgccggc caaccgggct 10080  
 cagctgaggt agatttgtca actataccaa ctagcatgat ctgcgaggt gagattgtaa 10140  
 ccggcggtgc ttcagcaatt tatggttcgg acgctgtatc aggtgttacc aacgttatcc 10200  
 ttaaagaaga ctttgaaggc tttgagttta acgcacgtac tagcggttct actgaaagtg 10260  
 taggcactca agagcactct tttgacattt tgggtggtgc aaacgttgca gatggacgtg 10320  
 gtaatgtaac cttctacgca gggtatgaac gtacaaaaga agtcatggct accgacattc 10380  
 gccaatcga tgcttgggga acaattaaaa acgaagccga tgggtggtgaa gatgatggta 10440  
 ttccagacag actacgtgta ccacgagttt attctgaaat gattaatgct accggtgtta 10500  
 tcaatgcatt tgggtggtgga attggtcgt caacctttga cagtaacggc aatcctattg 10560  
 cacaacaaga acgtgatggg actaacagct ttgcatttgg ttcattccct aatggctgtg 10620  
 acacatgttt caacactgaa gcatacgaaa actatatcc aggggtagaa agaataaacg 10680  
 ttggctcatc attcaacttt gattttaccg ataacattca attttacact gacttcagat 10740  
 atgtaaagtc agatattcag caacaatttc agccttcatt ccgttttggg aacattaata 10800  
 tcaatgttga agataacgcc tttttgaatg acgacttgcg tcagcaaagc ctcgatgcgg 10860  
 gtcaaaccga tgctagtttt gccaaagttt ttgatgaatt aggaaatcgc tcagcagaaa 10920  
 ataaacgcga acttttccgt tacgtaggtg gctttaaagg tggctttgat attagcgaaa 10980  
 ccatatttga ttacgacctt tactatgttt atggcgagac taataaccgt cgtaaaaccc 11040  
 ttaatgacct aattcctgat aactttgtcg cagctgtcga ctctgttatt gatcctgata 11100  
 ctggcttagc agcgtgtcgc tcacaagtag caagcgctca aggcgatgac tatacagatc 11160  
 ccgcgtctgt aaatggtagc gactgtgttg cttataaccc atttggcatg ggtcaagctt 11220  
 cagcagaagc ccgcgactgg gtttctgctg atgtgactcg tgaagacaaa ataactcaac 11280  
 aagtgattgg tggtagctctc ggtaccgatt ctgaagaact atttgagctt caaggtgggtg 11340  
 caatcgctat ggttggttgg tttgaatacc gtgaagaaac gtctggttca acaaccgatg 11400  
 aatttactaa agcaggttcc ttgacaagcg ctgcaacgcc agattcttat ggcgaatacg 11460  
 acgtgactga gtattttggt gaggtgaaca tcccagtact aaaagaatta cttttgcac 11520  
 atgagttgag ctttgacggt gcataccgta atgctgatta ctacatgcc ggtaagactg 11580  
 aagcatggaa agctgggtatg ttctactcac cattagagca acttgcatta cgtggtacgg 11640  
 taggtgaagc agtacgagca ccaaacattg cagaagcctt tagtccacgc tctcctgggt 11700  
 ttggcgcgt ttcagatcca tgtgatgcag ataacattaa tgacgatccg gatcgcggtg 11760  
 caaactgtgc agcattgggg atccctccag gattccaagc taatgataac gtcagtgtag 11820  
 ataccctatc tgggtgtaac ccagatctaa aacctgaaac atcaacatcc tttacaggtg 11880  
 gtcttggttg gacaccaacg tttgtgaca atctatcatt cactgtcgat tattatgata 11940  
 ttcaaattga ggatgctatt ttgtcagtag ccaccagac tgtggctgat aactgtgttg 12000  
 actcaactgg cggacctgac accgacttct gtagtcaagt tgatcgtaat ccaacgacct 12060  
 atgatattga acttggtcgc tctggttacc taaatgccgc ggcattgaat accaaaggta 12120  
 ttgaatttca agctgcatac tcattagatc tagagtcttt caacgcgcct ggtgaactac 12180  
 gcttcaacct attggggaac caattacttg aactagaacg tcttgaattc caaatcgctc 12240  
 ctgatgagat taatgatgaa aaaggcgaag taggtgatcc agagctgcag ttccgcctag 12300  
 gcacgatta ccgtctagat gatctaagt ttagctggaa cacgcgttat attgatagcg 12360  
 tagtaactta tgatgtctct gaaaatggtg gctctcctga agatttatat ccaggccaca 12420  
 taggctcaat gacaactcat gacttgagcg ctacatacta catcaatgag aacttcatga 12480  
 ttaacggtgg tgtacgtaac ctatttgacg cacttccacc tggatacact aacgatgcgc 12540  
 tatatgatct agttggtcgc cgtgcattcc taggtattaa ggtaatgatg taattaatta 12600  
 ttacgcctct aactaataaa aatgcaatct ctctgtagag attgcatttt tttatgaaat 12660

ccaatcttaa actggttctc cgagcatctt acgccttaaa aaccccgccc ctcaatgtaa 12720  
 cgccaaagt aattgcttac acgcacttac acaaacgaac aatttcatta acacgagaca 12780  
 cagctcacgc tttttatttt acccttgatt ttactacata aaattgcgtt ttagcgaca 12840  
 agtggtctcc caagctggtc gtatctgtaa ttattcagtc ccagggtgatt gtattgaccc 12900  
 ataagctcag gtagtctgct ctgccattag ctaaacaata ttgacaaaat ggcgataaaa 12960  
 tgtggcttag cgctaagttc accgtaagtt ttatcgcat taagtccca cagattatta 13020  
 acggaaaccc gctaaactga tggcaaaaat aaatagtga cacttgatg aagctactat 13080  
 tacttcgaat aagtgtacgc aaacagagac tgaggctcgg catagaaatg ccactacaac 13140  
 acctgagatg cgccgattca tacaagagtc ggatctcagt gttagccaac tgtctaaaat 13200  
 attaaatata agtgaagcta ccgtacgtaa gtggcgcaag cgtgactctg tcgaaaactg 13260  
 tcctaatacc ccgcaccatc tcaataccac gctaaccctt ttgcaagaat atgtggttgt 13320  
 gggcctgctg tatcaattga aaatgccatt agacagattg ctcaaagcaa cccaagagtt 13380  
 tatcaatcca aacgtgtcgc gctcaggttt agcaagatgt ttgaagcgtt atggcgtttc 13440  
 acgggtgagt gatattccaa gccacacgt accaatgcgc tactttaatc aaattccagt 13500  
 cactcaaggc agcgatgtgc aaacctacac cctgcactat gaaacgctgg caaaaacctt 13560  
 agccttacct agtaccgatg gtgacaatgt ggtgcaagtg gtgtctctca ccattccacc 13620  
 aaagttaacc gaagaagcac ccagttcaat tttgctcggc attgatcctc atagcgactg 13680  
 gatctatctc gacatatacc aagatggcaa tacacaagcc acgaatagat atatggctta 13740  
 tgtgctaaaa cacgggcatc tccatttacg aaagttaact gtgcgtaact atcacacctt 13800  
 tttacagcgc tttcctggag cgacgcaaaa tcgccgcccc tctaaagata tgcctgaaac 13860  
 aatcaacaag acgcctgaaa cacaggcacc cagtggagac tcataatgag ccagacctct 13920  
 aaacctacaa actcagcaac tgagcaagca caagactcac aagctgactc tcgtttaaat 13980  
 aaacgactaa aagatatgcc aattgctatt gttggcatgg cgagtatttt tgcaaaactct 14040  
 cgctatttga ataagttttg ggacttaatc agcgaaaaaa ttgatgcgat tactgaatta 14100  
 ccatcaactc actggcagcc tgaagaatat tacgacgcag ataaaaccgc agcagacaaa 14160  
 agctactgta aacgtggtgg ctttttgcca gatgtagact tcaaccaat ggagtttggc 14220  
 ctgccgcaa acattttgga actgaccgat tcacgcgaac tattatcact catcgttgct 14280  
 aaagaagtgt tggctgatgc taacttacct gagaattacg accgcgataa aattggtatc 14340  
 accttaggtg tcggcggtgg tcaaaaaatt agccacagcc taacagcgcg tctgcaatac 14400  
 ccagtattga agaaagtatt cgccaatagc ggcattagt acaccgacag cgaaatgctt 14460  
 atcaagaaat tccaagacca atatgtacac tgggaagaaa actcgttccc aggttcactt 14520  
 ggtaacgtta ttgcgggccc tatcgccaac cgcttcgatt ttggcgcat gaactgtgtg 14580  
 gttgatgctg cctgtgctgg atcacttgct gctatgcgta tggcgctaac agagctaact 14640  
 gaaggtcgct ctgaaatgat gatcaccgtt ggtgtgtgta ctgataactc acctctatg 14700  
 tatatgagct tttcaaaaac gccgccttt accactaacg aaaccattca gccatttgat 14760  
 atcgactcaa aaggcatgat gattggtgaa ggtattggca tgggtggcgt aaagcgtctt 14820  
 gaagatgcag agcgcgatgg cgaccgcatt tactctgtaa ttaaagggtt ggggtgcatc 14880  
 tctgacggtg agtttaaatc aatctatgcc cctcgcccat caggccaagc taaagcactt 14940  
 aaccgtgcct atgatgacgc aggttttgcg ccgcatacct taggtctaata tgaagctcac 15000  
 ggaacaggta ctgcagcagg tgacgcggca gagtttgccg gcctttgctc agtatttgct 15060  
 gaaggcaacg ataccaagca acacattgcg ctagggtcag ttaaatcaca aattggtcat 15120  
 actaaatcaa ctgcaggtac agcaggttta attaaagctg ctcttgcttt gcatcacaag 15180  
 gtactgccgc cgaccattaa cgtagtcag ccaagcccta aacttgatat cgaaaactca 15240  
 ccgttttata taaactga gactcgtcca tggttaccac gtgttgatgg tacgcgcgc 15300  
 cgcgcggtta ttagctcatt tggttttggt ggactaact tccattttgt actagaagag 15360  
 tacaaccaag aacacagccg tactgatagc gaaaaagcta agtatcgtca acgccaagt 15420  
 gcgcaaagct tccttgtag cgcaagcgat aaagcatcgc taattaacga gttaaacgta 15480  
 ctacgacgat ctgcaagcca agctgagttt atcctcaaag atgcagcagc aaactatggc 15540

gtacgtgagc ttgataaaaa tgcaccacgg atcgggttag ttgcaaacac agctgaagag 15600  
ttagcaggcc taattaagca agcacttgcc aaactagcag ctacgatga taacgcattg 15660  
cacctacctg gtggcactag ctaccgccc gctgcagtag aaggtaaagt tgccgcactg 15720  
tttgctggcc aagggtcaca atatctcaat atgggcccgtg accttacttg ttattacca 15780  
gagatgcgtc agcaatttgt aactgcagat aaagtatttg ccgcaaatga taaaacgccg 15840  
ttatcgcaaa ctctgtatcc aaagcctgta ttaataaag atgaattaaa ggctcaagaa 15900  
gccattttga ccaataccgc caatgcccaa agcgcaattg gtgcgatttc aatgggtcaa 15960  
tacgatttgt ttactgcggc tggctttaat gccgacatgg ttgcaggcca tagctttggg 16020  
gagctaagtg cactgtgtgc tgcaggtgtt atttcagctg atgactacta caagctggct 16080  
tttgctcgtg gtgaggctat ggcaacaaaa gcaccggcta aagacggcgt tgaagcagat 16140  
gcaggagcaa tgtttgcaat cataaccaag agtgctgcag accttgaaac cgttgaagcc 16200  
accatcgcta aatttgatgg ggtgaaagtc gctaactata acgcgccaac gcaatcagta 16260  
attgcaggcc caacagcaac taccgctgat gcgctaag cgctaactga gcttggttac 16320  
aaagcgatta acctgccagt atcaggtgca ttccacactg aacttggttg tcacgctcaa 16380  
gcgccatttg ctaaagcgat tgacgcagcc aaatttacta aaacaagccg agcactttac 16440  
tcaaatgcaa ctggcggact ttatgaaagc actgctgcaa agattaaagc ctgctttaag 16500  
aaacatatgc ttcaatcagt gcgctttact agccagctag aagccatgta caacgcggc 16560  
gcccgtgtat ttgttgaatt tggccaaaag aacatcttac aaaaattagt tcaaggcacg 16620  
cttgtcaaca ctgaaaatga agtttgact atctctatca accctaacc taaagttgat 16680  
agtgtatctg agcttaagca agcagcaatg cagctagcgg ttactggtgt ggtactcagt 16740  
gaaattgacc cataccaagc cgatattgcc gcaccagcga aaaagtcgcc aatgagcatt 16800  
tcgcttaatg ctgctaacca tatcagcaaa gcaactcgcg ctaagatggc caagtcttta 16860  
gagacaggta tcgtcacctc gcaaatagaa catgttattg aagaaaaaat cgttgaagtt 16920  
gagaaactgg ttgaagtcga aaagatcgtc gaaaaagtgg ttgaagtaga gaaagttgtt 16980  
gaggttgaag ctctgttaa ttacgtgcaa gccaatgcaa ttcaaaccg ttcagttgtc 17040  
gtccagtaa tagagaacca agtcgtgtct aaaaacagta agccagcagt ccagagcatt 17100  
agtgtgatg cactcagcaa cttttttgct gcacagcagc aaaccgcaca gttgcatcag 17160  
cagttcttag ctattccgca gcaatatggt gagacgttca ctacgctgat gaccgagcaa 17220  
gctaaactgg caagttctgg tgttgcaatt ccagagagtc tgcaacgctc aatggagcaa 17280  
ttccaccaac tacaagcgca aacactacaa agccacaccc agttccttga gatgcaagcg 17340  
ggtagcaaca ttgcagcgtt aaacctactc aatagcagcc aagcaactta cgtccagcc 17400  
attcacaatg aagcgattca aagccaagt gttcaaagcc aaactgcagt ccagccagta 17460  
atttcaacac aagttaacca tgtgtcagag cagccaactc aagctccagc tccaaaagcg 17520  
cagccagcac ctgtgacaac tgcagttcaa actgctccgg cacaagttgt tcgtcaagcc 17580  
gcaccagttc aagccgctat tgaaccgatt aatacaagt ttgcgactac aacgccttca 17640  
gccttcagcg ccgaaacagc cctgagcgca acaaaagtcc aagccactat gcttgaagtg 17700  
gttctgaga aaaccggtta cccaactgaa atgctagagc ttgaaatgga tatggaagcc 17760  
gatttaggca tcgattctat caagcgtgta gaaattcttg gcacagtaca agatgagcta 17820  
ccgggtctac ctgagcttag ccctgaagat ctactgagt gtcgaacgct aggcgaaatc 17880  
gttgactata tgggcagtaa actgccggct gaaggctcta tgaattctca gctgtctaca 17940  
ggttccgca ctgcgactcc tgcagcgaat ggtctttctg cggagaaagt tcaagcgact 18000  
atgatgtctg tggttgccga aaagactggc tacccaactg aaatgctaga gcttgaaatg 18060  
gatatggaag ccgatttagg catagattct atcaagcgcg ttgaaattct tggcacagta 18120  
caagatgagc taccgggtct acctgagctt agccctgaag atctagctga gtgtcgtact 18180  
ctaggcgaaa tcgttgacta tatgaactct aaactcgtg acggctctaa gctgccggct 18240  
gaaggctcta tgaattctca gctgtctaca agtgccgcag ctgcgactcc tgcagcgaat 18300  
ggtctctctg cggagaaagt tcaagcgact atgatgtctg tggttgccga aaagactggc 18360  
tacccaactg aaatgctaga acttgaaatg gatatggaag ctgaccttg catcgattca 18420

atcaagcgcg ttgaaattct tggcacagta caagatgagc taccgggttt acctgagcta 18480  
 aatccagaag atttggcaga gtgtcgtact cttggcgaaa tcgtgactta tatgaactct 18540  
 aaactcgctg acggctctaa gctgccagct gaaggctcta tgcactatca gctgtctaca 18600  
 agtaccgctg ctgcgactcc tgtagcgaat ggtctctctg cagaaaaagt tcaagcgacc 18660  
 atgatgtctg tagttgcaga taaaactggc tacccaactg aaatgcttga acttgaaatg 18720  
 gatattggaag ccgatttagg tatcgattct atcaagcgcg ttgaaattct tggcacagta 18780  
 caagatgagc taccgggttt acctgagcta aatccagaag atctagcaga gtgtcgcacc 18840  
 ctaggcgaaa tcgttgacta tatgggcagt aaactgccgg ctgaaggctc tgctaataca 18900  
 agtgccgctg cgtctcttaa tgttagtgcc gttgcggcgc ctcaagctgc tgcgactcct 18960  
 gtatcgaacg gtctctctgc agagaaagt caaagcacta tgatgtcagt agttgcagaa 19020  
 aagaccggct acccaactga aatgctagaa cttggcatgg atatggaagc cgatttaggt 19080  
 atcgactcaa ttaaacgcgt tgagattctt ggcacagtac aagatgagct accgggtcta 19140  
 ccagagctta atcctgaaga tttagctgag tgccgtacgc tgggcgaaat cgttgactat 19200  
 atgaactcta agctggctga cggctctaa cttccagctg aaggctctgc taatacaagt 19260  
 gccactgctg cgactcctgc agtgaatggt ctttctgctg acaaggtaca ggcgactatg 19320  
 atgtctgtag ttgctgaaaa gaccggctac ccaactgaaa tgctagaact tggcatggat 19380  
 atggaagcag accttggtat tgattctatt aagcgcgttg aaattcttg cagagtacaa 19440  
 gatgagctcc cagggtttacc tgagcttaat cctgaagatc tcgctgagtg ccgcacgctt 19500  
 ggcgaaatcg ttagctatat gaactctcaa ctggctgatg gctctaaact ttctacaagt 19560  
 gcggctgaag gctctgctga tacaagtgt gcaaatgctg caaagccggc agcaatttcg 19620  
 gcagaaccaa gtgttgagct tcctcctcat agcgaggtag cgctaaaaaa gcttaatgcy 19680  
 gcgaacaagc tagaaaattg tttcgccgca gacgcaagt ttgtgattaa cgatgatggt 19740  
 cacaacgcag gcgttttagc tgagaaactt attaaacaag gcctaaaagt agccgttgtg 19800  
 cgtttaccga aaggtcagcc tcaatcgcca ctttcaagcg atgttgctag ctttgagctt 19860  
 gcctcaagcc aagaatctga gcttgaagcc agtatcactg cagttatcgc gcagattgaa 19920  
 actcaggttg gcgctattgg tggctttatt cacttgcaac cagaagcgaa tacagaagag 19980  
 caaacggcag taaacctaga tgcgcaaagt tttactcac ttagcaatgc gttcttgtyg 20040  
 gccaaattat tgcaaccaa gctcgttgct ggagcagatg cgcgtcgtg ttttgtaaca 20100  
 gtaagccgta tcgacggtgg ctttggttac ctaaatactg acgccctaaa agatgctgag 20160  
 ctaaaccaag cagcattagc tggtttaact aaaaccttaa gccatgaatg gccacaagt 20220  
 ttctgtcgcg cgctagatat tgcaacagat gttgatgcaa cccatcttgc tgatgcaatc 20280  
 accagtgaac tatttgatag ccaagctcag ctacctgaag tgggcttaag cttaattgat 20340  
 ggcaaagtta accgcgtaac tctagttgct gctgaagctg cagataaaac agcaaaagca 20400  
 gagcttaaca gcacagataa aatcttagtg actggtgggg caaaaggggt gacatttgaa 20460  
 tgtgactggt cattagcatc tcgcagccag tctcacttta tcttagctgg gcgcagttaa 20520  
 ttacaagctt taccaagctg ggctgagggg aagcaaaacta gcgagctaaa atcagctgca 20580  
 atcgcacata ttatttctac tgggtcaaaag ccaacgccta agcaagttga agccgctgtg 20640  
 tggccagtgc aaagcagcat tgaaattaat gccgccttag ccgcctttaa caaagttggc 20700  
 gcctcagctg aatacgtcag catggatgtt accgatagcg ccgcaatcac agcagcactt 20760  
 aatggctcgt caaatgagat caccggtctt attcatggcg caggtgtact agccgacaag 20820  
 catattcaag acaagactct tgctgaactt gctaaagttt atggcactaa agtcaacggc 20880  
 ctaaaagcgc tgctcgcggc acttgagcca agcaaaatta aattacttgc tatgttctca 20940  
 tctgcagcag gtttttacgg taatatcggc caaagcgatt acgcgatgtc gaacgatatt 21000  
 cttaacaagg cagcgtgca gttcaccgct cgcaaccac aagctaaagt catgagcttt 21060  
 aactggggtc cttgggatgg cggcatggtt aaccagcgc ttaaaaagat gtttaccgag 21120  
 cgtggtgtgt acgttattcc actaaaagca ggtgcagagc tatttgccac tcagctattg 21180  
 gctgaaactg gcgtgcagtt gctcattggt acgtcaatgc aaggtggcag cgacactaaa 21240  
 gcaactgaga ctgcttctgt aaaaaagctt aatgcgggtg aggtgctaag tgcacgcat 21300

ccgcgtgctg gtgcacaaaa aacaccacta caagctgtca ctgcaacgcg tctgttaacc 21360  
ccaagtgccca tggcttcat tgaagatcac cgcattggcg gtaacagtgt gttgccaacg 21420  
gtatgcgccca tcgactggat gcgtgaagcg gcaagcgaca tgcttggcgc tcaagttaag 21480  
gtacttgatt acaagctatt aaaaggcatt gtatttgaga ctgatgagcc gcaagagtta 21540  
acacttgagc taacgccaga cgattcagac gaagctacgc tacaagcatt aatcagctgt 21600  
aatgggcgctc cgcaatacaa ggcgacgctt atcagtgata atgccgatat taagcaactt 21660  
aacaagcagt ttgatttaag cgctaaggcg attaccacag caaaagagct ttatagcaac 21720  
ggcaccttgt tccacggctc gcgtctacaa gggatccaat ctgtagtga gttcgtgat 21780  
caaggcttaa ttgctaaagt cgctctgcct aagggtgaac ttagcgattg tggtaggttc 21840  
ttgccgaaa cccacatggg ttggcagcaa ccttttgctg aggacttgct attacaagct 21900  
atgctggttt gggctcgct taaaactggc tcggcaagtt tgccatcaag cattggtgag 21960  
tttacctcat accaaccaat ggcctttggt gaaactggta ccatagagct tgaagtgatt 22020  
aagcacaaca aacgctcact tgaagcgaat gttgcgctat atcgtgacaa cggcgagtta 22080  
agtgccatgt ttaagtcagc taaaatcacc attagcaaaa gcttaaattc agcattttta 22140  
cctgctgtct tagcaaacga cagtgaaggcg aattagtga acaaacgcct aaagctagt 22200  
cgatgccgct gcgcacgca cttatcttac tgccaacacc gcagtttgaa gttactctg 22260  
tcgaccagtc agtattagcc agctatcaaa cactgcagcc tgagctaaat gccctgctta 22320  
atagtgcgcc gacacctgaa atgctcagca tcaactatctc agatgatagc gatgcaaaaca 22380  
gctttgagtc gcagctaaat gctgcgacca acgcaattaa caatggctat atcgtcaagc 22440  
ttgctacggc aactcacgct ttgttaatgc tgccctgcatt aaaagcggcg caaatgcgga 22500  
tccatctca tgcgcagctt gccgctatgc agcaagctaa atcgacgcca atgagtcaag 22560  
tatctggtga gctaaagctt ggcgctaag cgctaagcct agctcagact aatgcgctgt 22620  
ctcatgcttt aagccaagcc aagcgtaact taactgatgt cagcgtgaat gagtgttttg 22680  
agaacctcaa aagtgaacag cagttcacag aggtttattc gcttattcag caacttgcta 22740  
gccgcacca tgtgagaaaa gaggttaatc aagggttgga acttgccct aaacaagcca 22800  
aaagccacta ttggtttagc gaatttcacc aaaaccgtgt tgctgccatc aactttatta 22860  
atggccaaca agcaaccagc tatgtgctta ctcaaggctc aggttggtta gctgcgaaat 22920  
caatgctaaa ccagcaaaga ttaatgttta tcttgccggg taacagtcag caacaaataa 22980  
ccgcatcaat aactcagtta atgcagcaat tagagcgttt gcaggtaact gaggttaatg 23040  
agctttctct agaatgccaa cttagagctgc tcagcataat gtatgacaac ttagtcaacg 23100  
cagacaaact cactactcgc gatagtaagc ccgcttatca ggctgtgatt caagcaagct 23160  
ctgtagcgc tgcaaagcaa gagttaagcg cgcttaacga tgcactcaca gcgctgtttg 23220  
ctgagcaaac aaacgccaca tcaacgaata aaggcttaat ccaatacaaa acaccggcg 23280  
gcagttactt aaccctaaca ccgcttgcca gcaacaatga caacgcccga gcgggtcttg 23340  
ctttgtgcta tccgggtgtg ggaacggttt acgccgatat gcttaatgag ctgcatcagt 23400  
acttccctgc gctttacgcc aaacttgagc gtgaaggcga tttaaaggcg atgctacaag 23460  
cagaagatat ctatcatctt gaccctaaac atgctgcca aatgagctta ggtgacttag 23520  
ccattgctgg cgtggggagc agctacctgt taactcagct gctcaccgat gagttaata 23580  
ttaagcctaa ttttgcatca ggttactcaa tgggtgaagc atcaatgtgg gcaagcttag 23640  
gcgtatggca aaaccgcgat gcgctgatca gcaaaaccca aaccgaccg ctatttactt 23700  
ctgctatttc cggcaaattg accgcggtta gacaagcttg gcagcttgat gataccgag 23760  
cggaaatcca gtggaatagc tttgtggtta gaagtgaagc agcgcgatt gaagccttg 23820  
taaaagatta cccacacgct tacctcgca ttattcaagg ggatacctgc gtaatcgctg 23880  
gctgtgaaat ccaatgtaaa gcgctacttg cagcactggg taaacgcggt attgcagcta 23940  
atcgtgtaac ggcgatgcat acgcagcctg cgatgcaaga gcatcaaat gtgatggatt 24000  
tttatctgca accgttaaaa gcagagcttc ctagtgaat aagctttatc agcgcgctg 24060  
atttaactgc caagcaaagc gtgagtgagc aagcacttag cagccaagtc gttgctcagt 24120  
ctattgccga caccttctgc caaaccttg actttaccgc gctagtacat cagcccaac 24180



atcaaggcgc taagctgttt gttgaaattg gcgcggatag acaaaactgc accttgatag 24240  
 acaagattgt taaacaagat ggtgccagca gtgtacaaca tcaaccttgt tgcacagtgc 24300  
 ctatgaacgc aaaaggtagc caagatattt ccagcgtgat taaagcgctt ggccaattaa 24360  
 ttagccatca ggtgccatta tcggtgcaac catttattga tggactcaag cgcgagctaa 24420  
 cactttgcc aattgaccagc caacagctgg cagcacatgc aaatgttgac agcaagtttg 24480  
 agtctaacca agaccattta cttcaagggg aagtctaatt tcattaccag acaatgcttc 24540  
 taaccacctt tctgccaacc agaaaaggcg atctcaggca agtaaaacca gtaagcaaag 24600  
 caaaatcgcc attgtcgggt tagccactct gtatccagac gctaaaaccc cgcaagaatt 24660  
 ttggcagaat ttgctggata aacgcgactc tcgcagcacc ttaactaacg aaaaactcgg 24720  
 cgctaacagc caagattatc aaggtgtgca aggccaatct gaccgttttt attgtaataa 24780  
 aggcggtctac attgagaact tcagctttta tgctgcaggc taaaaattgc cggagcaaag 24840  
 cttaaatggc ttggacgaca gcttcctttg ggcgtcgtat actagccgta acgcactaat 24900  
 tgatgctggt attgatatca acggcgctga ttttaagccg gcaggtgtag tcatgggctg 24960  
 gctgtcgttc ccaactaccc gctcaaacga tctgtttttg ccaatttatc acagcgccgt 25020  
 tgaaaaagcc ctgcaagata aactaggcgt aaaggcattt aagctaagcc caactaatgc 25080  
 tcataccgct cgcgcggaat atgagagcag cctaaatgca gccaatggtg ccattgcca 25140  
 taacagctca aaagtgggtg ccgatgcact tggccttggc ggcgcacaac taagcctaga 25200  
 tgctgcctgt gctagtctcg tttactcatt aaagcttgcc tgcgattacc taagcactgg 25260  
 caaagccgat atcatgctag caggcgctgt atctggcgcg gatcctttct ttattaatat 25320  
 gggattctca atcttccacg cctaccaga ccatggtatc tcagtaccgt ttgatgccag 25380  
 cagtaaaggt ttgtttgctg gcgaaggcgc tggcgatta gtgcttaaac gtcttgaaga 25440  
 tgecgagcgc gacaatgaca aaatctatgc ggtgttagc ggcgtaggtc tatcaaacga 25500  
 cggtaaaggc cagtttgtat taagccctaa tccaaaaggt caggtgaagg cctttgaacg 25560  
 tgcttatgct gccagtga ttagaccaaa agacattgaa gtgattgagt gccacgcaac 25620  
 aggcacaccg cttggcgata aaattgagct cacttcaatg gaaaccttct ttgaagacaa 25680  
 gctgcaaggc accgatgcac cgtaattgg ctcagctaag tctaacttag gccacctatt 25740  
 aactgcagcg catgcgggga tcatgaagat gatcttcgcc atgaaagaag gttacctgcc 25800  
 gccaaagtac aatattagt atgctatcgc ttcgccgaaa aaactcttcg gtaaaccaac 25860  
 cctgcctagc atggttcaag gctggccaga taagccatcg aataatcatt ttggtgtaag 25920  
 aaccgcgtac gcaggcgtat cgggtatttg ctttgggtggc tgtaacgccc atctgttgct 25980  
 tgagtcatac aacggcaaa gaacagtaaa ggcagaagcc actcaagta cgcgtcaagc 26040  
 tgagcgcta aaagtgggtg gccttgctc gcactttggg cctcttagca gcattaatgc 26100  
 actcaacaat gctgtgacct aagatgggaa tggctttatc gaactgccga aaaagcgctg 26160  
 gaaaaggcctt gaaaagcaca gtgaactgtt agctgaattt ggcttagcat ctgcgcaaaa 26220  
 aggtgcttat gttgataact tcgagctgga ctttttacgc tttaaactgc cgccaaacga 26280  
 agatgaccgt ttgatctcac agcagctaatt gctaattgca gtaacagacg aagccattcg 26340  
 tgatgccaag cttgagccgg ggcaaaaagt agctgtatta gtggcaatgg aaactgagct 26400  
 tgaactgat cagttccgcg gccgggttaa cttgcatact caattagcgc aaagtcttg 26460  
 cgccatgggc tgaggtttat caacgatga ataccaagcg cttgaagcca tcgccatgga 26520  
 cagcgtgctt gatgctgcca agctcaatca gtacaccagc tttattggta atattatggc 26580  
 gtcacgcgtg gcgtcactat gggactttta tggcccagcc ttcactattt cagcagcaga 26640  
 gcaatctgtg agccgctgta tcgatgtggc gcaaaacctc atcatggagg ataacctaga 26700  
 tgcggtgggt attgcagcg tcgatctctc tggtagcttt gagcaagtca ttttataaaa 26760  
 tgccattgca cctgtagcca ttgagccaaa cctcgaagca agccttaatc caacatcagc 26820  
 aagctggaat gtcggtgaag gtgctggcgc ggtcgtgctt gttaaaaatg aagctacatc 26880  
 gggctgctca tacggccaaa ttgatgcact tggctttgct aaaactgccg aaacagcgtt 26940  
 ggctaccgac aagctactga gccaaactgc cacagacttt aataagggtt aagtattgga 27000  
 aactatggca gcgcctgcta gccaaattca attagcgcca atagttagct ctcaagtgc 27060

tcacactgct gcagagcagc gtgttggtca ctgctttgct gcagcgggta tggcaagcct 27120  
 attacacggc ttacttaact taaatactgt agcccaaacc aataaagcca attgcgcgct 27180  
 tatcaacaat atcagtgaat accaattatc acagctgttg attagccaaa cagcgagcga 27240  
 acaacaagca ttaaccgcgc gtttaagcaa tgagcttaaa tccgatgcta aacaccaact 27300  
 ggtaagcaa gtcaccttag gtggcgtga tatctaccag catattgttg atacaccgct 27360  
 tgcaagcctt gaaagcatta ctcaaaaatt ggcgcaagc acagcatcga cagtgggtcaa 27420  
 ccaagttaaa cctattaagg ccgctggctc agtcgaaatg gctaactcat tcgaaacgga 27480  
 aagctcagca gagccacaaa taacaattgc agcacaacag actgcaaaca ttggcgtcac 27540  
 cgctcaggca accaaacgtg aattaggtac cccaccaatg acaacaaata ccattgctaa 27600  
 tacagcaaat aatttagaca agactcttga gactgttgct ggcaatactg ttgctagcaa 27660  
 gggtggctct ggcgacatag tcaattttca acagaaccaa caattggctc aacaagctca 27720  
 cctgcgcttt cttgaaagcc gcagtgcggg tatgaaggtg gctgatgctt tattgaagca 27780  
 acagctagct caagtaacag gccaaactat cgataatcag gccctcgata ctcaagccgt 27840  
 cgatactcaa acaagcgaga atgtagcgat tgccgcagaa tcaccagttc aagttacaac 27900  
 acctgttcaa gttacaacac ctgttcaa atcaggtgtg gagttaaaac cagatcacgc 27960  
 taatgtgcc ccatcacgc cgccagtgc tgcattaaag ccgtgtatct ggaactatgc 28020  
 cgatttagtt gagtacgcag aaggcgatat cgccaaggta tttggcagt attatgcat 28080  
 tatcgacagc tactcgcgcc gcgtacgtct accgaccact gactacctgt tggatcgcg 28140  
 cgtagcaaaa cttgatgcga ccatcaatca atttaagcca tgctcaatga ccactgagta 28200  
 cgacatccct gttgatgcgc cgtacttagt agacggacaa atcccttggg cggtagcagt 28260  
 agaatcaggc caatgtgact tgatgcttat tagctatctc ggtatcgact ttgagaacaa 28320  
 aggcgagcgg gtttatcgac tactcgattg taccctcacc ttcctaggcg acttgccacg 28380  
 tggcgagat accctacgtt acgacattaa gatcaataac tatgctcgca acggcgacac 28440  
 cdtgtgttc tcttctcgt atgagtgtt tgttgcgac aagatgatcc tcaagatgga 28500  
 tggcggtgc gctggcttct tcaactgatga agagcttgcc gacggtaaaag gcgtgattcg 28560  
 cacagaagaa gagattaaag ctgcgagcct agtgcaaaaag caacgcttta atccgttact 28620  
 agattgtcct aaaacccaat ttagttatgg tgatattcat aagctattaa ctgctgatat 28680  
 tgagggttgt tttggcccaa gccacagtgg cgtccaccag ccgtcacttt gtttcgcatc 28740  
 tgaaaaattc ttgatgattg aacaagtcag caaggttgat cgcactggcg gtacttgggg 28800  
 acttggtta attgagggtc ataagcagct tgaagcagac cactggtact tcccatgtca 28860  
 tttcaagggc gaccaagtga tggctggctc gctaattggc gaaggttggt gccagttatt 28920  
 gcagttctat atgctgcacc ttggtatgca taccacaaact aaaaatggc gtttccaacc 28980  
 tcttgaaaac gcctcacagc aagtacgctg tcgcggtaaa gtgctgccac aatcaggcgt 29040  
 gctaacttac cgtatggaag tgactgaaat cggtttcagt ccacgccc atgctaaagc 29100  
 taacatcgat atcttgctta atggcaaaag ggtagtggat ttccaaaacc taggggtgat 29160  
 gataaaagag gaagatgagt gtactcgta tccacttttg actgaatcaa caacggctag 29220  
 cactgcacaa gtaaacgctc aaacaagtgc gaaaaaggta tacaagccag catcagtcga 29280  
 tgcgccatta atggcacaaa ttcctgatct gactaaagag ccaacaagg gcgttattcc 29340  
 gatttcccat gttgaagcac caattacgcc agactacccg aaccgtgtac ctgatacagt 29400  
 gccattcacg ccgtatcaca tgtttgagtt tgctacaggc aatatcgaaa actgtttcgg 29460  
 gccagagttc tcaatctatc gcggcatgat cccaccacgt acaccatgcg gtgacttaca 29520  
 agtgaccaca cgtgtgattg aagttaacgg taagcgtggc gactttaaaa agccatcatc 29580  
 gtgtatcgct gaatatgaag tgctgcaga tgcgtggtat ttcgataaaa acagccacgg 29640  
 cgagtgatg ccatattcaa ttttaattga gatctcactg caacctaacg gctttatctc 29700  
 aggttacatg ggcacaaccc taggttccc tggccttgag ctgttcttcc gtaacttaga 29760  
 cggtagcggg gagttactac gtgaagtaga ttacgtggg aaaaccatcc gtaacgactc 29820  
 acgtttatta tcaacagtga tggccggcac taacatcatc caaagcttta gcttcgagct 29880  
 aagcactgac ggtgagcctt tctatcgcg cactgcggta tttggctatt ttaaaggtag 29940

cgcacttaaa gatcagctag gcctagataa cggtaaagtc actcagccat ggcattgtagc 30000  
 taacggcggt gctgcaagca ctaagggtgaa cctgcttgat aagagctgcc gtcacttttaa 30060  
 tgcgccagct aaccagccac actatcgtct agccgggtggt cagctgaact ttatcgacag 30120  
 tgttgaaatt gttgataatg gcggcaccga aggttttaggt tacttgtagt ccgagcgcac 30180  
 cattgaccca agtgattggt tcttccagtt ccacttccac caagatccgg ttatgccagg 30240  
 ctcttaggt gttgaagcaa ttattgaaac catgcaagct tacgctatta gtaaagactt 30300  
 gggcgagat ttcaaaaatc ctaagtttgg tcagatttta tcgaacatca agtggaaagta 30360  
 tcgcggtcaa atcaatccgc tgaacaagca gatgtctatg gatgtcagca ttacttcaat 30420  
 caaagatgaa gacggttaaga aagtcacac aggtaatgcc agcttgagta aagatggtct 30480  
 gcgcataac gaggtcttcg atatactat cagcatcgaa gaatctgtat aaatcgaggt 30540  
 gactgtctgg ctattttact caatttctgt gtcaaaagt ctcacctata ttcatagggt 30600  
 gcgcgctttt ttctggaaat tgagcaaaag tatctgcgtc ctaactcgat ttataagaat 30660  
 gggttaattg aaaagaacaa cagctaagag ccgcaagctc aatataaata attagggtc 30720  
 ttacaaataa tgaatcctac agcaactaac gaaatgcttt ctccgtggcc atgggctgtg 30780  
 acagagtc aaatcagttt tgacgtgcaa gtgatggaac aacaacttaa agatttttagc 30840  
 cgggcatggt acgtgggtcaa tcatgccgac cagggctttg gtattgcgca aactgccgat 30900  
 atcgtgactg aacaagcggc aaacagcaca gatttacctg ttagtgcttt tactcctgca 30960  
 ttaggtaccg aaagcctagg cgacaataat ttccgcccgc ttacggcgt taaatacgct 31020  
 tattacgcag gcgctatggc aaacgggtatt tcatctgaag agctagtgtat tgccttaggt 31080  
 caagctggca tttgtgtgg ttctgttggg gcagccggtc ttattccaag tcgctgtgaa 31140  
 gcggcaatta accgtattca agcagcgtg ccaaatggcc cttatatgtt taaccttacc 31200  
 catagtccca gcgagccagc attagagcgt ggcagcgtag agctattttt aaagcataag 31260  
 gtacgcaccg ttgaagcacc agctttctta ggtctaacac cacaatcgt ctattaccgt 31320  
 gcagcaggat tgagccgaga cgcacaaggt aaagtgtgtg ttggtaacaa ggttatcgct 31380  
 aaagtaagtc gcaccgaagt ggctgaaaag tttatgatgc cagcgcccgc aaaaatgcta 31440  
 caaaaactag ttgatgacgg ttcaattacc gctgagcaaa tggagctggc gcaacttgta 31500  
 cctatggctg acgacatcac tgcagaggcc gattcaggtg gccatactga taaccgtcca 31560  
 ttagtaacat tgctgccaac catttttagc ctgaaagaag aaattcaagc taaataccaa 31620  
 tacgacactc ctattcgtgt cggttgtggt ggcggtgtgg gtacgcctga tgcagcgtg 31680  
 gcaacgttta acatggcgcc ggcttatatt gttaccggct ctatcaacca agcttgtgtt 31740  
 gaagcgggag caagtgatca cactcgtaaa ttacttgcca ccaactgaaat ggccgatgtg 31800  
 actatggcac cagctgcaga tatgttcgag atgggcgtaa aactgcaggt ggttaagcgc 31860  
 ggcacgctat tcccaatgcg cgctaacaag ctatatgaga tctacaccg ttacgattca 31920  
 atcgaagcga tcccattaga cgagcgtgaa aagcttgaga aacaagtatt ccgctcaagc 31980  
 ctatgtaaaa tatgggcagg tacagtggcg cactttaacg agcgcgaccc taagcaaatc 32040  
 gaacgcgcag agggtaaccc taagcgtaaa atggcattga tttccgttg gtacttaggt 32100  
 ctttctagtc gctgggtcaa ctcaggcgaa gtgggtcgtg aaatggatta tcaaatttgg 32160  
 gctggccctg ctctcggtgc atttaaccaa tgggcaaaag gcagttactt agataactat 32220  
 caagaccgaa atgccgtcga tttggcaaaag cacttaatgt acggcgcgcc ttacttaaat 32280  
 cgtattaact cgctaacggc tcaaggcggt aaagtgccag cacagttact tcgctggaag 32340  
 ccaaaccaaa gaatggccta atacacttac aaagcaccag tctaaaaagc cactaatctt 32400  
 gattagtggc tttttttatt gtgttcaata tgaggctatt tagcctgtaa gcctgaaaat 32460  
 atcagcactc tgactttaca agcaattat aattaaggca gggtctact catttatact 32520  
 gctagcaaac aagcaagttg ccagtaaaaa caacaaggta cctgatttat atcgtcataa 32580  
 aagttggcta gagattcgtt attgatcttt actgattaga gtcgctctgt ttggaaaaag 32640  
 gtttctcgtt atcatcaaaa tacactctca aacctttaat caattacaac ttaggctttc 32700  
 tgcgggcatt tttatcttat ttgccacagc tgtatttgcc ttaggtttt ggggtgcaact 32760  
 accattaatt gaggcctcat tagttaaat atctgagcaa gagctcacct ctttaaat 32820

cgcttttcag caaatgagaa agccactaca aaccattaat tacgactatg cggtgtggga 32880  
 cagaacctac agctatatga aatcaaactc agcgagcgct aaaaggtact atgaaaaaca 32940  
 tgagtacca gatgatacgt tcaagagttt aaaagtcgac ggagtattta tattcaaccg 33000  
 tacaaatcag ccagttttta gtaaaggttt taatcataga aatgatatac cgctgggtctt 33060  
 tgaattaact gactttaaac aacatccaca aaacatcgca ttatctccac aaaccaaaaca 33120  
 ggcacacca ccggaagta agccgttaga cccccctgat gatgtgcctt ctacccatgg 33180  
 ggttatcgcc acacgatacg gtccagcaat ttatagctct accagcattt taaaatctga 33240  
 tcgtagcggc tcccaacttg gttatttagt cttcattagg ttaattgatg aatggttcat 33300  
 cgctgagcta tcgcaataca ctgccgcagg tgttgaaatc gctatggctg atgccgcaga 33360  
 cgcacaatta gcgagattag gcgcaaacac taagcttaat aaagtaaccg ctacatccga 33420  
 acggttaata actaatgtcg atggtaagcc tctgttgaag ttagtgcttt accataccaa 33480  
 taaccaaccg ccgccgatgc tagattacag tataataatt ctattagttg agatgtcatt 33540  
 ttactgatc ctgcgttatt tcctttactc ctacttctta gtcaggccag ttagaaagct 33600  
 ggcttcagat attaaaaaaa tggataaaaag tcgtgaaatt aaaaagctaa ggtatcacta 33660  
 ccctattact gagctagtca aagttgcgac tcacttcaac gccctaattg ggacgattca 33720  
 ggaacaaact aaacagctta atgaacaagt ttttattgat aaattaacca atattcccaa 33780  
 tcgtcgcgct tttgagcagc gacttgaaac ctattgcaa ctgctagccc ggcaacaaat 33840  
 tggctttact ctcatcattg ccgatgtgga tcatttttaa gagtacaacg atactcttg 33900  
 gcaccttgct ggggatgaag cattaataaa agtggcaca acactatcgc aacagtttta 33960  
 ccgtgcagaa gatatttggt cccgttttgg tggatgaaga tttattatgt tatttcgaga 34020  
 catacctgat gagcccttgc agagaaagct cgatgcgatg ctgcaactct ttgcagagct 34080  
 caactctact catccaaact catcaaccgc taattacgtt actgtgagcc ttgggggttg 34140  
 cacagttggt gctgttgatg attttgaatt taaaagtga tcgcatatta ttggcagtc 34200  
 ggctgcatta atcgagata aggcgcttta tcatgctaaa gcctgtggtc gtaaccagtt 34260  
 gtcaaaaact actattactg ttgatgatg tgagcaatta gaagcaata aaatcgggtca 34320  
 tcaagcctaa actcgttcga gtactttccc ctaagtcaga gctatttgcc acttcaagat 34380  
 gtggctacaa ggcttactct ttcaaaacct gcataatag aacacagcaa aatacaataa 34440  
 ttttaagtcaa tttagcctat taaacagagt taatgacagc tcatggctgc aacttattag 34500  
 ctatttctag caatataaaa acttatccat tagtagtaac caataaaaaa actaatatat 34560  
 aaaactatth aatcattatt ttacagatga ttagtaccac ccacacttaa gctggctata 34620  
 ttgcactag taaaaataaa cattagatcg ggttcagatc aatttacgag tctcgataa 34680  
 aatgtacaat aattcactta atttaatact gcatattttt acaagtagag agcggtgatg 34740  
 aaacaaaata cgaaaggctt tacattaatt gaattagtca tcgtgattat tattctcgg 34800  
 atacttgctg ctgtggcact gccgaaattc atcaatgttc aagatgacgc taggatctct 34860  
 gcgatgagcg gtcagttttc atcatttgaa agtgccgtaa aactatacca tagcgggttg 34920  
 ttagccaaag gtcacaacac tgcgggtgaa aagctctcag gctttggcca aggtaatgtt 34980  
 gcatcaagtg acacagggtt tccgtactca acatcaggca cgagtactga tgtgcataaa 35040  
 gcttgtggtg aactatggca tggcattacc gatacagact tcacaattgg tgcggttagt 35100  
 gatggcgatc taatgactgc agatgtcgat attgcttaca cctatcgtgg tgatagtgt 35160  
 atctatcgcg atctgtattt tattcagcgc tcattaccta ctaaggatga gaactacaaa 35220  
 tttaaaactg gtgaaataga aattattgat gctttctaca accctgacgg ctcaactggt 35280  
 caattaccat aaatttggcg cttatctaa ttgtacttgc tctgaccgac acaataatg 35340  
 tcgtttctca gcatatatca aaatacacag caaaaatttg gggtagcta tatagctaac 35400  
 cccaaatcat atctaacttt aactgcatc taattccaaa cagtatccag ccaaagcct 35460  
 aaactattgt tgactcagcg ctaaaatatg cgatgcaaca aacaagtctt ggatcgcaat 35520  
 acctgagcta tcaaaaatgg tcacctatc agcactttga cgtcctgttg cggaactcgt 35580  
 tatcacctga ccaatctcaa ttatcggcgt atttctgcta tgttgaaact caccaataac 35640  
 aatagattga gaagcaaagt cgcaaaacaa gcgagcatga ctatataggt cagttggcaa 35700

ctcttgctta cccactttat cagcgcccat tgcagaaata tgcgttcctg cttgtaccca 35760  
 ctgcgcttca aataaaggcg cttgagctgt ggttgctgtg ataataatat ctgcttggtc 35820  
 acaagcagct tgtgcatcac aagcttcggc attaatgcct ttttctaata aacgcttaac 35880  
 caagttttca gttttgctag cactacggcc aactaccaat accttagtta atgaacgaac 35940  
 cttgctcact gctagcactt catattcagc ctgatgaccg gtacaaaaaa cagttaatac 36000  
 cgtagcatct tctctcgcga ggtaactcac tgctactgca tcggcagcac cagtgcggta 36060  
 agcattaacg gtagtggcag caatcaccgn ctgcaacata ccggttaatg gatcgagtaa 36120  
 aaatacgta gtgccgtggc atggtaaacc atgtttatgg ttatcaggcc aatagctgcc 36180  
 tgttttcag ccgacaaggt ttggcgttga agccgacttt aatgagaaca tttcattaag 36240  
 gttcgcgccc tgtgcattaa ctaccgggaa caaggttgct ttatcatcta cggcagcgac 36300  
 aaacgcttct ttaacagcga tataagccag ctcatgggag atgagctttg atgtttgcgc 36360  
 ttcagttaaa tagatcatat taccaccctt gactcagatt ccagatctca tagccaccat 36420  
 tatcaccatc agtatcaaat acatggtact gagcgtgcat tgaagctggt gcacaggcgt 36480  
 ggttcggcaa aatatgtaga cgactaccta ccgggaactg cgctaaatca ataacgccgc 36540  
 catcaactgc ttcaataatg ccgtgctctt gattaacagt tataacctgt agacctgata 36600  
 acacgtgacc gctgtcgtca cactactaac cataaccaca atcttttggc tgctctgcag 36660  
 tacctctatc acccgaaaga gccatccaac ccgcatcaat gaaaatccag tttttatcag 36720  
 gattatgacc aataacactg gtcactaccg ttgcggcaat atcagttaac tgacacacgt 36780  
 ttagccctgc catgactaaa tcgaagaagg tgtacacacc cgctctaacc tcggtgatcc 36840  
 catcaagggt ttgatagctt tgcgctgttg gtgttgaaac aataactaac atgtcacatt 36900  
 gcataccgcg tgcgcgaatg cgtcagcagc ttgtacagcc gctgcaactt catthttgcgc 36960  
 cgcatcaatt aattgctggt tttcaaaaca ttgatatgac tcaccagcgt gagtnagtac 37020  
 gccgtgaaaa ctgcgtgcgc cagacgttag tatctgagca atttcaatca acttatcggc 37080  
 ttccggtgga ataccaccac gatggccatc acaatcaatt tcaattaatg ctggtatttg 37140  
 gcagtcataa gaaccacaga aatgatttag ctgatgcgct tgctcaacac tatcaagtaa 37200  
 aactcttgca ttaatacctt ggtccaacat tttagcaata ccgcgcaact taccatcggc 37260  
 aatacctact gcataaataa tgtctgtgta accttagat gctaaggcct cggcctctt 37320  
 taccgttgat acagtgactg gtgagttttt agtgggtaat aaaaactcgg ctgcttcaag 37380  
 tgatcttaac gttttaaaat gcggtcttag gtttgacact aatccttcaa ttttttggcg 37440  
 tagttgactg aggttattaa taaatactgg cttattttaca tataaaaacg gtgtatcaat 37500  
 tgcttgatac tgactttgct ggtcgtgga aagtatttga gtagatggca tctttaatat 37560  
 cctagttcat caatcaatct aacaagtttg atgcctagcc acagtggctt gtattcatga 37620  
 tgctttggaa aatgcttata ttcaaagtat ttgaaagaca tcaaacttct tgtttaatgc 37680  
 tcagtatcca ccagcacgca tttattttat attaaactatt atcaagatat agattaggtt 37740  
 caaaccaaat gattagtact gaagatctac gttttatcag cgtaatcgcc agtcatcgca 37800  
 ccttagctga tgccgctaga aactaaata tcacgccacc atcagtgaca ttaagggttg 37860  
 agcatattga aaagaaacta tcgattagcc tgatc 37895

&lt;210&gt; 2

&lt;211&gt; 654

&lt;212&gt; PRT

<213> *Shewanella putrefaciens*

&lt;400&gt; 2

Met Lys Gln Thr Leu Met Ala Ile Ser Ile Met Ser Leu Phe Ser Phe

1

5

10

15

Asn Ala Leu Ala Ala Gln His Glu His Asp His Ile Thr Val Asp Tyr

20	25	30
Glu Gly Lys Ala Ala Thr Glu His Thr Ile Ala His Asn Gln Ala Val		
35	40	45
Ala Lys Thr Leu Asn Phe Ala Asp Thr Arg Ala Phe Glu Gln Ser Ser		
50	55	60
Lys Asn Leu Val Ala Lys Phe Asp Lys Ala Thr Ala Asp Ile Leu Arg		
65	70	75
Ala Glu Phe Ala Phe Ile Ser Asp Glu Ile Pro Asp Ser Val Asn Pro		
85	90	95
Ser Leu Tyr Arg Gln Ala Gln Leu Asn Met Val Pro Asn Gly Tyr Lys		
100	105	110
Val Ser Asp Gly Ile Tyr Gln Val Arg Gly Thr Asp Leu Ser Asn Leu		
115	120	125
Thr Leu Ile Arg Ser Asp Asn Gly Trp Ile Ala Tyr Asp Val Leu Leu		
130	135	140
Thr Lys Glu Ala Ala Lys Ala Ser Leu Gln Phe Ala Leu Lys Asn Leu		
145	150	155
Pro Lys Asp Gly Asp Pro Val Val Ala Met Ile Tyr Ser His Ser His		
165	170	175
Ala Asp His Phe Gly Gly Ala Arg Gly Val Gln Glu Met Phe Pro Asp		
180	185	190
Val Lys Val Tyr Gly Ser Asp Asn Ile Thr Lys Glu Ile Val Asp Glu		
195	200	205
Asn Val Leu Ala Gly Asn Ala Met Ser Arg Arg Ala Ala Tyr Gln Tyr		
210	215	220
Gly Ala Thr Leu Gly Lys His Asp His Gly Ile Val Asp Ala Ala Leu		
225	230	235
Gly Lys Gly Leu Ser Lys Gly Glu Ile Thr Tyr Val Ala Pro Asp Tyr		
245	250	255
Thr Leu Asn Ser Glu Gly Lys Trp Glu Thr Leu Thr Ile Asp Gly Leu		
260	265	270
Glu Met Val Phe Met Asp Ala Ser Gly Thr Glu Ala Glu Ser Glu Met		

16

530	535	540
Thr Leu Phe Asp Phe Leu Ala Val Lys Ile Asp Ser Gln Gln Ala Ala		
545	550	555 560
Lys His Gly Leu Val Lys Met Asn Val Ile Thr Pro Asp Thr Lys Asp		
565	570	575
Ile Leu Tyr Ile Glu Leu Ser Asn Gly Asn Leu Ser Asn Ala Val Val		
580	585	590
Asp Lys Glu Gln Leu Met Val Asn Lys Ala Asp Val Asn Arg Ile Leu		
595	600	605
Leu Gly Gln Val Thr Leu Lys Ala Leu Leu Ala Ser Gly Asp Ala Lys		
610	615	620
Leu Thr Gly Asp Lys Thr Ala Phe Ser Lys Ile Ala Asp Ser Met Val		
625	630	635 640
Glu Phe Thr Pro Asp Phe Glu Ile Val Pro Thr Pro Val Lys		
645	650	

<210> 3  
 <211> 277  
 <212> PRT  
 <213> *Shewanella putrefaciens*

<400> 3
Ser Thr Lys Ala Ser Ala Arg Val Val Ala Lys Phe Asn Val Glu Glu
1 5 10 15
Ala Ala Ile Ser Ile Gln Gln Cys Gln Gly Ile Ser Leu Ala Phe Arg
20 25 30
Tyr Ser Asp Asp Leu His Gly Leu Leu Cys His Trp Asn Asp Ala Ala
35 40 45
Asn Met Gln Gln Glu Lys Ala Glu Ile Leu Gly Leu Gly Ser Lys Gln
50 55 60
Pro Glu Ala Asn Pro Lys Asn Ser Ser Ser Glu Leu Leu Ala Leu Gly
65 70 75 80
Ile Asp Gln Lys Leu Leu Val Gln Arg Gln Asn Leu Gln His Glu Val
85 90 95



Lys His Asp Ala Ile Ala Asp Ser Ile Asp Val Cys His Ser Leu Ser  
 100 105 110

Lys Pro Ala Asn Val Gly Leu Phe Thr Glu Ser Leu Ala Ser Phe Asp  
 115 120 125

Phe Ala Phe Ser Lys Leu Ser Leu Ala Leu Gly Leu Gly Lys Ala Lys  
 130 135 140

Ile Tyr Ser Glu Lys Leu Ala Trp Leu Asp Phe Phe Arg Asp Arg Gln  
 145 150 155 160

Leu Ala Glu Pro Leu Ala Leu Leu Ala Arg Lys Glu Ser Glu Ser Phe  
 165 170 175

Tyr His Ser Leu Ile Ser His Ile Asn Thr Ser Asn Arg Cys Arg Glu  
 180 185 190

Ile Asp Val Gly Phe Glu Ile Ser Ala Ser Asp Thr Glu Glu Lys Ser  
 195 200 205

Ala Gln Ser Ala Gly Lys Asn Asp Ala Thr Cys Ile Gly Val Leu Leu  
 210 215 220

Trp Asp Gly Ser His Ser Val Asn Phe His Val Gly Thr Gln Ala Phe  
 225 230 235 240

Gln Ala Asp Ser Leu Arg Pro Lys Gly Lys Asp Gly Tyr Glu Phe Arg  
 245 250 255

Trp Glu Asn Pro Arg Ile Glu Ser His Gln Ser Leu Leu Ala Arg Leu  
 260 265 270

Tyr Gly Arg Val Met  
 275

<210> 4

<211> 1480

<212> DNA

<213> *Shewanella putrefaciens*

<400> 4

gctagtctta gctgasrthr ysaasragct cgaacaacag ctttaaaatt cacttcttct 60  
 gctgcaatac ttatttgctg acactgacca atactcagtg caaaacgata actatcatca 120  
 agatgggaaar gvavaaaysh asnvaggaaa asrgngncys gngysraaha rgtyrsrasa 180  
 shscccagta aacaatgccca attatcagca gcgttcattt gctgttcttt agcctcaatc 240  
 aaacctaacc cagacttttg tggctcagcg ttaggcttat taggycyshs trasnasaaa 300

```

aasnmtgn gn gysaaggygy srysgnrgaa asnrysasns raactcgact ctagtaaagc 360
aagaccaata tcttgtttta acaaaacctg tcgctgatta agttgatgct caaccttggtg 420
atccgcaata gcatcggaata tsrsrgaagy asgnysvagn arggnasngn hsgvayshsa 480
saaaaassra tcaacacaaat ggctcaagct tttaggtgca ttaactccaa gaaaagtttc 540
gctcagtgca gagaagtcaa acgcaaaaga ttttagcgat aatgccagca svacyshssr 600
srysraaasn vagyhthrgs raasrhasha ahsryssraa ccaagtcctt tcgctttaat 660
gtaagactcc ttgagcgccc acaaatcaaa aaagcggctc cgctgcaagg cctctggtaa 720
cgctaacaag gctcgctttt gygyysaays tyrsrgysaa trashharga sarggnaagr 780
aaaaargysg ctgattcaga gaaataatga ctaagaatag agtggatatt ggtgctgtta 840
cggcaacgct caatgtcgac gccaaactca atactagcag agtcagtttc srgsrhtyrh 900
ssrsrhasn thrsrasnar gcysarggas vagyhgsraa srasthrget ccttgcttgc 960
ctgactggcg cctttattat cagcagtgca aatgcctact aatagccaat ctccactatg 1020
actcacatta aagtggaccc cggtttgagy ssraagnsra agyysasnas aathrcysgy 1080
vatrasgyr hssrvaasnh hsvagythrg ngcaaattgc gcatcactca atctaggctt 1140
acctttgtcg ccatattcaa agcgccattc attggggcgt atttactat gttgtgacaa 1200
taaagcgcg aaahgnaaas srargrysgy ysasgytyrg hargtrgasn rarggsrhsg 1260
nsraaargaa tagcctctta ccattaaacc ttgagtttta gcttcttgtt taatgtagcg 1320
attaacctta attaaactcat cttcaggcag ccatgactta accaactcty rgyargvamt 1380
gygnthrysa aggnystyra rgasnvaysg asgrtrsrys vagtgtagtc tgggtatcgc 1440
actcttgat tgtaacgga cagaagtata aggaaatcaa 1480

```

&lt;210&gt; 5

&lt;211&gt; 970

&lt;212&gt; PRT

<213> *Shewanella putrefaciens*

&lt;400&gt; 5

```

Met Ser Met Phe Leu Asn Ser Lys Leu Ser Arg Ser Val Lys Leu Ala
  1             5             10             15

```

```

Ile Ser Ala Gly Leu Thr Ala Ser Leu Ala Met Pro Val Phe Ala Glu
      20             25             30

```

```

Glu Thr Ala Ala Glu Glu Gln Ile Glu Arg Val Ala Val Thr Gly Ser
      35             40             45

```

```

Arg Ile Ala Lys Ala Glu Leu Thr Gln Pro Ala Pro Val Val Ser Leu
      50             55             60

```

```

Ser Ala Glu Glu Leu Thr Lys Phe Gly Asn Gln Asp Leu Gly Ser Val
      65             70             75             80

```

```

Leu Ala Glu Leu Pro Ala Ile Gly Ala Thr Asn Thr Ile Ile Gly Asn
      85             90             95

```

```

Asn Asn Ser Asn Ser Ser Ala Gly Val Ser Ser Ala Asp Leu Arg Arg
      100            105            110

```

Leu Gly Ala Asn Arg Thr Leu Val Leu Val Asn Gly Lys Arg Tyr Val  
 115 120 125

Ala Gly Gln Pro Gly Ser Ala Glu Val Asp Leu Ser Thr Ile Pro Thr  
 130 135 140

Ser Met Ile Ser Arg Val Glu Ile Val Thr Gly Gly Ala Ser Ala Ile  
 145 150 155 160

Tyr Gly Ser Asp Ala Val Ser Gly Val Ile Asn Val Ile Leu Lys Glu  
 165 170 175

Asp Phe Glu Gly Phe Glu Phe Asn Ala Arg Thr Ser Gly Ser Thr Glu  
 180 185 190

Ser Val Gly Thr Gln Glu His Ser Phe Asp Ile Leu Gly Gly Ala Asn  
 195 200 205

Val Ala Asp Gly Arg Gly Asn Val Thr Phe Tyr Ala Gly Tyr Glu Arg  
 210 215 220

Thr Lys Glu Val Met Ala Thr Asp Ile Arg Gln Phe Asp Ala Trp Gly  
 225 230 235 240

Thr Ile Lys Asn Glu Ala Asp Gly Gly Glu Asp Asp Gly Ile Pro Asp  
 245 250 255

Arg Leu Arg Val Pro Arg Val Tyr Ser Glu Met Ile Asn Ala Thr Gly  
 260 265 270

Val Ile Asn Ala Phe Gly Gly Gly Ile Gly Arg Ser Thr Phe Asp Ser  
 275 280 285

Asn Gly Asn Pro Ile Ala Gln Gln Glu Arg Asp Gly Thr Asn Ser Phe  
 290 295 300

Ala Phe Gly Ser Phe Pro Asn Gly Cys Asp Thr Cys Phe Asn Thr Glu  
 305 310 315 320

Ala Tyr Glu Asn Tyr Ile Pro Gly Val Glu Arg Ile Asn Val Gly Ser  
 325 330 335

Ser Phe Asn Phe Asp Phe Thr Asp Asn Ile Gln Phe Tyr Thr Asp Phe  
 340 345 350

Arg Tyr Val Lys Ser Asp Ile Gln Gln Gln Phe Gln Pro Ser Phe Arg  
 355 360 365

Phe Gly Asn Ile Asn Ile Asn Val Glu Asp Asn Ala Phe Leu Asn Asp  
 370 375 380

Asp Leu Arg Gln Gln Met Leu Asp Ala Gly Gln Thr Asn Ala Ser Phe  
 385 390 395 400

Ala Lys Phe Phe Asp Glu Leu Gly Asn Arg Ser Ala Glu Asn Lys Arg  
 405 410 415

Glu Leu Phe Arg Tyr Val Gly Gly Phe Lys Gly Gly Phe Asp Ile Ser  
 420 425 430

Glu Thr Ile Phe Asp Tyr Asp Leu Tyr Tyr Val Tyr Gly Glu Thr Asn  
 435 440 445

Asn Arg Arg Lys Thr Leu Asn Asp Leu Ile Pro Asp Asn Phe Val Ala  
 450 455 460

Ala Val Asp Ser Val Ile Asp Pro Asp Thr Gly Leu Ala Ala Cys Arg  
 465 470 475 480

Ser Gln Val Ala Ser Ala Gln Gly Asp Asp Tyr Thr Asp Pro Ala Ser  
 485 490 495

Val Asn Gly Ser Asp Cys Val Ala Tyr Asn Pro Phe Gly Met Gly Gln  
 500 505 510

Ala Ser Ala Glu Ala Arg Asp Trp Val Ser Ala Asp Val Thr Arg Glu  
 515 520 525

Asp Lys Ile Thr Gln Gln Val Ile Gly Gly Thr Leu Gly Thr Asp Ser  
 530 535 540

Glu Glu Leu Phe Glu Leu Gln Gly Gly Ala Ile Ala Met Val Val Gly  
 545 550 555 560

Phe Glu Tyr Arg Glu Glu Thr Ser Gly Ser Thr Thr Asp Glu Phe Thr  
 565 570 575

Lys Ala Gly Phe Leu Thr Ser Ala Ala Thr Pro Asp Ser Tyr Gly Glu  
 580 585 590

Tyr Asp Val Thr Glu Tyr Phe Val Glu Val Asn Ile Pro Val Leu Lys  
 595 600 605

Glu Leu Pro Phe Ala His Glu Leu Ser Phe Asp Gly Ala Tyr Arg Asn  
 610 615 620

Ala Asp Tyr Ser His Ala Gly Lys Thr Glu Ala Trp Lys Ala Gly Met  
 625 630 635 640  
 Phe Tyr Ser Pro Leu Glu Gln Leu Ala Leu Arg Gly Thr Val Gly Glu  
 645 650 655  
 Ala Val Arg Ala Pro Asn Ile Ala Glu Ala Phe Ser Pro Arg Ser Pro  
 660 665 670  
 Gly Phe Gly Arg Val Ser Asp Pro Cys Asp Ala Asp Asn Ile Asn Asp  
 675 680 685  
 Asp Pro Asp Arg Val Ser Asn Cys Ala Ala Leu Gly Ile Pro Pro Gly  
 690 695 700  
 Phe Gln Ala Asn Asp Asn Val Ser Val Asp Thr Leu Ser Gly Gly Asn  
 705 710 715 720  
 Pro Asp Leu Lys Pro Glu Thr Ser Thr Ser Phe Thr Gly Gly Leu Val  
 725 730 735  
 Trp Thr Pro Thr Phe Ala Asp Asn Leu Ser Phe Thr Val Asp Tyr Tyr  
 740 745 750  
 Asp Ile Gln Ile Glu Asp Ala Ile Leu Ser Val Ala Thr Gln Thr Val  
 755 760 765  
 Ala Asp Asn Cys Val Asp Ser Thr Gly Gly Pro Asp Thr Asp Phe Cys  
 770 775 780  
 Ser Gln Val Asp Arg Asn Pro Thr Thr Tyr Asp Ile Glu Leu Val Arg  
 785 790 795 800  
 Ser Gly Tyr Leu Asn Ala Ala Ala Leu Asn Thr Lys Gly Ile Glu Phe  
 805 810 815  
 Gln Ala Ala Tyr Ser Leu Asp Leu Glu Ser Phe Asn Ala Pro Gly Glu  
 820 825 830  
 Leu Arg Phe Asn Leu Leu Gly Asn Gln Leu Leu Glu Leu Arg Leu  
 835 840 845  
 Glu Phe Gln Asn Arg Pro Asp Glu Ile Asn Asp Glu Lys Gly Glu Val  
 850 855 860  
 Gly Asp Pro Glu Leu Gln Phe Arg Leu Gly Ile Asp Tyr Arg Leu Asp  
 865 870 875 880



Lys Arg Tyr Gly Val Ser Arg Val Ser Asp Ile Gln Ser Pro His Val  
 130 135 140

Pro Met Arg Tyr Phe Asn Gln Ile Pro Val Thr Gln Gly Ser Asp Val  
 145 150 155 160

Gln Thr Tyr Thr Leu His Tyr Glu Thr Leu Ala Lys Thr Leu Ala Leu  
 165 170 175

Pro Ser Thr Asp Gly Asp Asn Val Val Gln Val Val Ser Leu Thr Ile  
 180 185 190

Pro Pro Lys Leu Thr Glu Glu Ala Pro Ser Ser Ile Leu Leu Gly Ile  
 195 200 205

Asp Pro His Ser Asp Trp Ile Tyr Leu Asp Ile Tyr Gln Asp Gly Asn  
 210 215 220

Thr Gln Ala Thr Asn Arg Tyr Met Ala Tyr Val Leu Lys His Gly Pro  
 225 230 235 240

Phe His Leu Arg Lys Leu Leu Val Arg Asn Tyr His Thr Phe Leu Gln  
 245 250 255

Arg Phe Pro Gly Ala Thr Gln Asn Arg Arg Pro Ser Lys Asp Met Pro  
 260 265 270

Glu Thr Ile Asn Lys Thr Pro Glu Thr Gln Ala Pro Ser Gly Asp Ser  
 275 280 285

<210> 7

<211> 2756

<212> PRT

<213> *Shewanella putrefaciens*

<400> 7

Met Ser Gln Thr Ser Lys Pro Thr Asn Ser Ala Thr Glu Gln Ala Gln  
 1 5 10 15

Asp Ser Gln Ala Asp Ser Arg Leu Asn Lys Arg Leu Lys Asp Met Pro  
 20 25 30

Ile Ala Ile Val Gly Met Ala Ser Ile Phe Ala Asn Ser Arg Tyr Leu

35	40	45	
Asn Lys Phe Trp Asp Leu Ile Ser Glu Lys Ile Asp Ala Ile Thr Glu			
50	55	60	
Leu Pro Ser Thr His Trp Gln Pro Glu Glu Tyr Tyr Asp Ala Asp Lys			
65	70	75	80
Thr Ala Ala Asp Lys Ser Tyr Cys Lys Arg Gly Gly Phe Leu Pro Asp			
85	90	95	
Val Asp Phe Asn Pro Met Glu Phe Gly Leu Pro Pro Asn Ile Leu Glu			
100	105	110	
Leu Thr Asp Ser Ser Gln Leu Leu Ser Leu Ile Val Ala Lys Glu Val			
115	120	125	
Leu Ala Asp Ala Asn Leu Pro Glu Asn Tyr Asp Arg Asp Lys Ile Gly			
130	135	140	
Ile Thr Leu Gly Val Gly Gly Gly Gln Lys Ile Ser His Ser Leu Thr			
145	150	155	160
Ala Arg Leu Gln Tyr Pro Val Leu Lys Lys Val Phe Ala Asn Ser Gly			
165	170	175	
Ile Ser Asp Thr Asp Ser Glu Met Leu Ile Lys Lys Phe Gln Asp Gln			
180	185	190	
Tyr Val His Trp Glu Glu Asn Ser Phe Pro Gly Ser Leu Gly Asn Val			
195	200	205	
Ile Ala Gly Arg Ile Ala Asn Arg Phe Asp Phe Gly Gly Met Asn Cys			
210	215	220	
Val Val Asp Ala Ala Cys Ala Gly Ser Leu Ala Ala Met Arg Met Ala			
225	230	235	240
Leu Thr Glu Leu Thr Glu Gly Arg Ser Glu Met Met Ile Thr Gly Gly			
245	250	255	
Val Cys Thr Asp Asn Ser Pro Ser Met Tyr Met Ser Phe Ser Lys Thr			
260	265	270	
Pro Ala Phe Thr Thr Asn Glu Thr Ile Gln Pro Phe Asp Ile Asp Ser			
275	280	285	
Lys Gly Met Met Ile Gly Glu Gly Ile Gly Met Val Ala Leu Lys Arg			



290	295	300
Leu Glu Asp Ala Glu Arg Asp Gly Asp Arg Ile Tyr Ser Val Ile Lys		
305	310	315 320
Gly Val Gly Ala Ser Ser Asp Gly Lys Phe Lys Ser Ile Tyr Ala Pro		
	325	330 335
Arg Pro Ser Gly Gln Ala Lys Ala Leu Asn Arg Ala Tyr Asp Asp Ala		
	340	345 350
Gly Phe Ala Pro His Thr Leu Gly Leu Ile Glu Ala His Gly Thr Gly		
	355	360 365
Thr Ala Ala Gly Asp Ala Ala Glu Phe Ala Gly Leu Cys Ser Val Phe		
	370	375 380
Ala Glu Gly Asn Asp Thr Lys Gln His Ile Ala Leu Gly Ser Val Lys		
	385	390 395 400
Ser Gln Ile Gly His Thr Lys Ser Thr Ala Gly Thr Ala Gly Leu Ile		
	405	410 415
Lys Ala Ala Leu Ala Leu His His Lys Val Leu Pro Pro Thr Ile Asn		
	420	425 430
Val Ser Gln Pro Ser Pro Lys Leu Asp Ile Glu Asn Ser Pro Phe Tyr		
	435	440 445
Leu Asn Thr Glu Thr Arg Pro Trp Leu Pro Arg Val Asp Gly Thr Pro		
	450	455 460
Arg Arg Ala Gly Ile Ser Ser Phe Gly Phe Gly Gly Thr Asn Phe His		
	465	470 475 480
Phe Val Leu Glu Glu Tyr Asn Gln Glu His Ser Arg Thr Asp Ser Glu		
	485	490 495
Lys Ala Lys Tyr Arg Gln Arg Gln Val Ala Gln Ser Phe Leu Val Ser		
	500	505 510
Ala Ser Asp Lys Ala Ser Leu Ile Asn Glu Leu Asn Val Leu Ala Ala		
	515	520 525
Ser Ala Ser Gln Ala Glu Phe Ile Leu Lys Asp Ala Ala Ala Asn Tyr		
	530	535 540
Gly Val Arg Glu Leu Asp Lys Asn Ala Pro Arg Ile Gly Leu Val Ala		

545	550	555	560
Asn Thr Ala Glu Glu Leu Ala Gly Leu Ile Lys Gln Ala Leu Ala Lys			
565	570	575	
Leu Ala Ala Ser Asp Asp Asn Ala Trp Gln Leu Pro Gly Gly Thr Ser			
580	585	590	
Tyr Arg Ala Ala Ala Val Glu Gly Lys Val Ala Ala Leu Phe Ala Gly			
595	600	605	
Gln Gly Ser Gln Tyr Leu Asn Met Gly Arg Asp Leu Thr Cys Tyr Tyr			
610	615	620	
Pro Glu Met Arg Gln Gln Phe Val Thr Ala Asp Lys Val Phe Ala Ala			
625	630	635	640
Asn Asp Lys Thr Pro Leu Ser Gln Thr Leu Tyr Pro Lys Pro Val Phe			
645	650	655	
Asn Lys Asp Glu Leu Lys Ala Gln Glu Ala Ile Leu Thr Asn Thr Ala			
660	665	670	
Asn Ala Gln Ser Ala Ile Gly Ala Ile Ser Met Gly Gln Tyr Asp Leu			
675	680	685	
Phe Thr Ala Ala Gly Phe Asn Ala Asp Met Val Ala Gly His Ser Phe			
690	695	700	
Gly Glu Leu Ser Ala Leu Cys Ala Ala Gly Val Ile Ser Ala Asp Asp			
705	710	715	720
Tyr Tyr Lys Leu Ala Phe Ala Arg Gly Glu Ala Met Ala Thr Lys Ala			
725	730	735	
Pro Ala Lys Asp Gly Val Glu Ala Asp Ala Gly Ala Met Phe Ala Ile			
740	745	750	
Ile Thr Lys Ser Ala Ala Asp Leu Glu Thr Val Glu Ala Thr Ile Ala			
755	760	765	
Lys Phe Asp Gly Val Lys Val Ala Asn Tyr Asn Ala Pro Thr Gln Ser			
770	775	780	
Val Ile Ala Gly Pro Thr Ala Thr Thr Ala Asp Ala Ala Lys Ala Leu			
785	790	795	800
Thr Glu Leu Gly Tyr Lys Ala Ile Asn Leu Pro Val Ser Gly Ala Phe			

805	810	815
His Thr Glu Leu Val Gly His Ala Gln Ala Pro Phe Ala Lys Ala Ile 820	825	830
Asp Ala Ala Lys Phe Thr Lys Thr Ser Arg Ala Leu Tyr Ser Asn Ala 835	840	845
Thr Gly Gly Leu Tyr Glu Ser Thr Ala Ala Lys Ile Lys Ala Ser Phe 850	855	860
Lys Lys His Met Leu Gln Ser Val Arg Phe Thr Ser Gln Leu Glu Ala 865	870	875 880
Met Tyr Asn Asp Gly Ala Arg Val Phe Val Glu Phe Gly Pro Lys Asn 885	890	895
Ile Leu Gln Lys Leu Val Gln Gly Thr Leu Val Asn Thr Glu Asn Glu 900	905	910
Val Cys Thr Ile Ser Ile Asn Pro Asn Pro Lys Val Asp Ser Asp Leu 915	920	925
Gln Leu Lys Gln Ala Ala Met Gln Leu Ala Val Thr Gly Val Val Leu 930	935	940
Ser Glu Ile Asp Pro Tyr Gln Ala Asp Ile Ala Ala Pro Ala Lys Lys 945	950	955 960
Ser Pro Met Ser Ile Ser Leu Asn Ala Ala Asn His Ile Ser Lys Ala 965	970	975
Thr Arg Ala Lys Met Ala Lys Ser Leu Glu Thr Gly Ile Val Thr Ser 980	985	990
Gln Ile Glu His Val Ile Glu Glu Lys Ile Val Glu Val Glu Lys Leu 995	1000	1005
Val Glu Val Glu Lys Ile Val Glu Lys Val Val Glu Val Glu Lys Val 1010	1015	1020
Val Glu Val Glu Ala Pro Val Asn Ser Val Gln Ala Asn Ala Ile Gln 1025	1030	1035 1040
Thr Arg Ser Val Val Ala Pro Val Ile Glu Asn Gln Val Val Ser Lys 1045	1050	1055
Asn Ser Lys Pro Ala Val Gln Ser Ile Ser Gly Asp Ala Leu Ser Asn		

1060	1065	1070
Phe Phe Ala Ala Gln Gln Gln Thr Ala Gln Leu His Gln Gln Phe Leu		
1075	1080	1085
Ala Ile Pro Gln Gln Tyr Gly Glu Thr Phe Thr Thr Leu Met Thr Glu		
1090	1095	1100
Gln Ala Lys Leu Ala Ser Ser Gly Val Ala Ile Pro Glu Ser Leu Gln		
1105	1110	1115 1120
Arg Ser Met Glu Gln Phe His Gln Leu Gln Ala Gln Thr Leu Gln Ser		
1125	1130	1135
His Thr Gln Phe Leu Glu Met Gln Ala Gly Ser Asn Ile Ala Ala Leu		
1140	1145	1150
Asn Leu Leu Asn Ser Ser Gln Ala Thr Tyr Ala Pro Ala Ile His Asn		
1155	1160	1165
Glu Ala Ile Gln Ser Gln Val Val Gln Ser Gln Thr Ala Val Gln Pro		
1170	1175	1180
Val Ile Ser Thr Gln Val Asn His Val Ser Glu Gln Pro Thr Gln Ala		
1185	1190	1195 1200
Pro Ala Pro Lys Ala Gln Pro Ala Pro Val Thr Thr Ala Val Gln Thr		
1205	1210	1215
Ala Pro Ala Gln Val Val Arg Gln Ala Ala Pro Val Gln Ala Ala Ile		
1220	1225	1230
Glu Pro Ile Asn Thr Ser Val Ala Thr Thr Thr Pro Ser Ala Phe Ser		
1235	1240	1245
Ala Glu Thr Ala Leu Ser Ala Thr Lys Val Gln Ala Thr Met Leu Glu		
1250	1255	1260
Val Val Ala Glu Lys Thr Gly Tyr Pro Thr Glu Met Leu Glu Leu Glu		
1265	1270	1275 1280
Met Asp Met Glu Ala Asp Leu Gly Ile Asp Ser Ile Lys Arg Val Glu		
1285	1290	1295
Ile Leu Gly Thr Val Gln Asp Glu Leu Pro Gly Leu Pro Glu Leu Ser		
1300	1305	1310
Pro Glu Asp Leu Ala Glu Cys Arg Thr Leu Gly Glu Ile Val Asp Tyr		

1315	1320	1325
Met Gly Ser Lys Leu Pro Ala Glu Gly Ser Met Asn Ser Gln Leu Ser		
1330	1335	1340
Thr Gly Ser Ala Ala Ala Thr Pro Ala Ala Asn Gly Leu Ser Ala Glu		
1345	1350	1355 1360
Lys Val Gln Ala Thr Met Met Ser Val Val Ala Glu Lys Thr Gly Tyr		
1365	1370	1375
Pro Thr Glu Met Leu Glu Leu Glu Met Asp Met Glu Ala Asp Leu Gly		
1380	1385	1390
Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Gly Thr Val Gln Asp Glu		
1395	1400	1405
Leu Pro Gly Leu Pro Glu Leu Ser Pro Glu Asp Leu Ala Glu Cys Arg		
1410	1415	1420
Thr Leu Gly Glu Ile Val Asp Tyr Met Asn Ser Lys Leu Ala Asp Gly		
1425	1430	1435 1440
Ser Lys Leu Pro Ala Glu Gly Ser Met Asn Ser Gln Leu Ser Thr Ser		
1445	1450	1455
Ala Ala Ala Ala Thr Pro Ala Ala Asn Gly Leu Ser Ala Glu Lys Val		
1460	1465	1470
Gln Ala Thr Met Met Ser Val Val Ala Glu Lys Thr Gly Tyr Pro Thr		
1475	1480	1485
Glu Met Leu Glu Leu Glu Met Asp Met Glu Ala Asp Leu Gly Ile Asp		
1490	1495	1500
Ser Ile Lys Arg Val Glu Ile Leu Gly Thr Val Gln Asp Glu Leu Pro		
1505	1510	1515 1520
Gly Leu Pro Glu Leu Asn Pro Glu Asp Leu Ala Glu Cys Arg Thr Leu		
1525	1530	1535
Gly Glu Ile Val Thr Tyr Met Asn Ser Lys Leu Ala Asp Gly Ser Lys		
1540	1545	1550
Leu Pro Ala Glu Gly Ser Met His Tyr Gln Leu Ser Thr Ser Thr Ala		
1555	1560	1565
Ala Ala Thr Pro Val Ala Asn Gly Leu Ser Ala Glu Lys Val Gln Ala		

1570	1575	1580	
Thr Met Met Ser Val Val Ala Asp Lys Thr Gly Tyr Pro Thr Glu Met			
1585	1590	1595	1600
Leu Glu Leu Glu Met Asp Met Glu Ala Asp Leu Gly Ile Asp Ser Ile			
1605	1610	1615	
Lys Arg Val Glu Ile Leu Gly Thr Val Gln Asp Glu Leu Pro Gly Leu			
1620	1625	1630	
Pro Glu Leu Asn Pro Glu Asp Leu Ala Glu Cys Arg Thr Leu Gly Glu			
1635	1640	1645	
Ile Val Asp Tyr Met Gly Ser Lys Leu Pro Ala Glu Gly Ser Ala Asn			
1650	1655	1660	
Thr Ser Ala Ala Ala Ser Leu Asn Val Ser Ala Val Ala Ala Pro Gln			
1665	1670	1675	1680
Ala Ala Ala Thr Pro Val Ser Asn Gly Leu Ser Ala Glu Lys Val Gln			
1685	1690	1695	
Ser Thr Met Met Ser Val Val Ala Glu Lys Thr Gly Tyr Pro Thr Glu			
1700	1705	1710	
Met Leu Glu Leu Gly Met Asp Met Glu Ala Asp Leu Gly Ile Asp Ser			
1715	1720	1725	
Ile Lys Arg Val Glu Ile Leu Gly Thr Val Gln Asp Glu Leu Pro Gly			
1730	1735	1740	
Leu Pro Glu Leu Asn Pro Glu Asp Leu Ala Glu Cys Arg Thr Leu Gly			
1745	1750	1755	1760
Glu Ile Val Asp Tyr Met Asn Ser Lys Leu Ala Asp Gly Ser Lys Leu			
1765	1770	1775	
Pro Ala Glu Gly Ser Ala Asn Thr Ser Ala Thr Ala Ala Thr Pro Ala			
1780	1785	1790	
Val Asn Gly Leu Ser Ala Asp Lys Val Gln Ala Thr Met Met Ser Val			
1795	1800	1805	
Val Ala Glu Lys Thr Gly Tyr Pro Thr Glu Met Leu Glu Leu Gly Met			
1810	1815	1820	
Asp Met Glu Ala Asp Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile			

1825	1830	1835	1840
Leu Gly Thr Val Gln Asp Glu Leu Pro Gly Leu Pro Glu Leu Asn Pro			
1845	1850	1855	
Glu Asp Leu Ala Glu Cys Arg Thr Leu Gly Glu Ile Val Ser Tyr Met			
1860	1865	1870	
Asn Ser Gln Leu Ala Asp Gly Ser Lys Leu Ser Thr Ser Ala Ala Glu			
1875	1880	1885	
Gly Ser Ala Asp Thr Ser Ala Ala Asn Ala Ala Lys Pro Ala Ala Ile			
1890	1895	1900	
Ser Ala Glu Pro Ser Val Glu Leu Pro Pro His Ser Glu Val Ala Leu			
1905	1910	1915	1920
Lys Lys Leu Asn Ala Ala Asn Lys Leu Glu Asn Cys Phe Ala Ala Asp			
1925	1930	1935	
Ala Ser Val Val Ile Asn Asp Asp Gly His Asn Ala Gly Val Leu Ala			
1940	1945	1950	
Glu Lys Leu Ile Lys Gln Gly Leu Lys Val Ala Val Val Arg Leu Pro			
1955	1960	1965	
Lys Gly Gln Pro Gln Ser Pro Leu Ser Ser Asp Val Ala Ser Phe Glu			
1970	1975	1980	
Leu Ala Ser Ser Gln Glu Ser Glu Leu Glu Ala Ser Ile Thr Ala Val			
1985	1990	1995	2000
Ile Ala Gln Ile Glu Thr Gln Val Gly Ala Ile Gly Gly Phe Ile His			
2005	2010	2015	
Leu Gln Pro Glu Ala Asn Thr Glu Glu Gln Thr Ala Val Asn Leu Asp			
2020	2025	2030	
Ala Gln Ser Phe Thr His Val Ser Asn Ala Phe Leu Trp Ala Lys Leu			
2035	2040	2045	
Leu Gln Pro Lys Leu Val Ala Gly Ala Asp Ala Arg Arg Cys Phe Val			
2050	2055	2060	
Thr Val Ser Arg Ile Asp Gly Gly Phe Gly Tyr Leu Asn Thr Asp Ala			
2065	2070	2075	2080
Leu Lys Asp Ala Glu Leu Asn Gln Ala Ala Leu Ala Gly Leu Thr Lys			

2085	2090	2095
Thr Leu Ser His Glu Trp Pro Gln Val Phe Cys Arg Ala Leu Asp Ile		
2100	2105	2110
Ala Thr Asp Val Asp Ala Thr His Leu Ala Asp Ala Ile Thr Ser Glu		
2115	2120	2125
Leu Phe Asp Ser Gln Ala Gln Leu Pro Glu Val Gly Leu Ser Leu Ile		
2130	2135	2140
Asp Gly Lys Val Asn Arg Val Thr Leu Val Ala Ala Glu Ala Ala Asp		
2145	2150	2155 2160
Lys Thr Ala Lys Ala Glu Leu Asn Ser Thr Asp Lys Ile Leu Val Thr		
2165	2170	2175
Gly Gly Ala Lys Gly Val Thr Phe Glu Cys Ala Leu Ala Leu Ala Ser		
2180	2185	2190
Arg Ser Gln Ser His Phe Ile Leu Ala Gly Arg Ser Glu Leu Gln Ala		
2195	2200	2205
Leu Pro Ser Trp Ala Glu Gly Lys Gln Thr Ser Glu Leu Lys Ser Ala		
2210	2215	2220
Ala Ile Ala His Ile Ile Ser Thr Gly Gln Lys Pro Thr Pro Lys Gln		
2225	2230	2235 2240
Val Glu Ala Ala Val Trp Pro Val Gln Ser Ser Ile Glu Ile Asn Ala		
2245	2250	2255
Ala Leu Ala Ala Phe Asn Lys Val Gly Ala Ser Ala Glu Tyr Val Ser		
2260	2265	2270
Met Asp Val Thr Asp Ser Ala Ala Ile Thr Ala Ala Leu Asn Gly Arg		
2275	2280	2285
Ser Asn Glu Ile Thr Gly Leu Ile His Gly Ala Gly Val Leu Ala Asp		
2290	2295	2300
Lys His Ile Gln Asp Lys Thr Leu Ala Glu Leu Ala Lys Val Tyr Gly		
2305	2310	2315 2320
Thr Lys Val Asn Gly Leu Lys Ala Leu Leu Ala Ala Leu Glu Pro Ser		
2325	2330	2335
Lys Ile Lys Leu Leu Ala Met Phe Ser Ser Ala Ala Gly Phe Tyr Gly		



2340	2345	2350
Asn Ile Gly Gln Ser Asp Tyr Ala Met Ser Asn Asp Ile Leu Asn Lys		
2355	2360	2365
Ala Ala Leu Gln Phe Thr Ala Arg Asn Pro Gln Ala Lys Val Met Ser		
2370	2375	2380
Phe Asn Trp Gly Pro Trp Asp Gly Gly Met Val Asn Pro Ala Leu Lys		
2385	2390	2395
Lys Met Phe Thr Glu Arg Gly Val Tyr Val Ile Pro Leu Lys Ala Gly		
2405	2410	2415
Ala Glu Leu Phe Ala Thr Gln Leu Leu Ala Glu Thr Gly Val Gln Leu		
2420	2425	2430
Leu Ile Gly Thr Ser Met Gln Gly Gly Ser Asp Thr Lys Ala Thr Glu		
2435	2440	2445
Thr Ala Ser Val Lys Lys Leu Asn Ala Gly Glu Val Leu Ser Ala Ser		
2450	2455	2460
His Pro Arg Ala Gly Ala Gln Lys Thr Pro Leu Gln Ala Val Thr Ala		
2465	2470	2475
Thr Arg Leu Leu Thr Pro Ser Ala Met Val Phe Ile Glu Asp His Arg		
2485	2490	2495
Ile Gly Gly Asn Ser Val Leu Pro Thr Val Cys Ala Ile Asp Trp Met		
2500	2505	2510
Arg Glu Ala Ala Ser Asp Met Leu Gly Ala Gln Val Lys Val Leu Asp		
2515	2520	2525
Tyr Lys Leu Leu Lys Gly Ile Val Phe Glu Thr Asp Glu Pro Gln Glu		
2530	2535	2540
Leu Thr Leu Glu Leu Thr Pro Asp Asp Ser Asp Glu Ala Thr Leu Gln		
2545	2550	2555
Ala Leu Ile Ser Cys Asn Gly Arg Pro Gln Tyr Lys Ala Thr Leu Ile		
2565	2570	2575
Ser Asp Asn Ala Asp Ile Lys Gln Leu Asn Lys Gln Phe Asp Leu Ser		
2580	2585	2590
Ala Lys Ala Ile Thr Thr Ala Lys Glu Leu Tyr Ser Asn Gly Thr Leu		

2595                      2600                      2605  
 Phe His Gly Pro Arg Leu Gln Gly Ile Gln Ser Val Val Gln Phe Asp  
 2610                      2615                      2620  
 Asp Gln Gly Leu Ile Ala Lys Val Ala Leu Pro Lys Val Glu Leu Ser  
 2625                      2630                      2635                      2640  
 Asp Cys Gly Glu Phe Leu Pro Gln Thr His Met Gly Gly Ser Gln Pro  
 2645                      2650                      2655  
 Phe Ala Glu Asp Leu Leu Leu Gln Ala Met Leu Val Trp Ala Arg Leu  
 2660                      2665                      2670  
 Lys Thr Gly Ser Ala Ser Leu Pro Ser Ser Ile Gly Glu Phe Thr Ser  
 2675                      2680                      2685  
 Tyr Gln Pro Met Ala Phe Gly Glu Thr Gly Thr Ile Glu Leu Glu Val  
 2690                      2695                      2700  
 Ile Lys His Asn Lys Arg Ser Leu Glu Ala Asn Val Ala Leu Tyr Arg  
 2705                      2710                      2715                      2720  
 Asp Asn Gly Glu Leu Ser Ala Met Phe Lys Ser Ala Lys Ile Thr Ile  
 2725                      2730                      2735  
 Ser Lys Ser Leu Asn Ser Ala Phe Leu Pro Ala Val Leu Ala Asn Asp  
 2740                      2745                      2750  
 Ser Glu Ala Asn  
 2755

&lt;210&gt; 8

&lt;211&gt; 771

&lt;212&gt; PRT

<213> *Shewanella putrefaciens*

&lt;400&gt; 8

Met Pro Leu Arg Ile Ala Leu Ile Leu Leu Pro Thr Pro Gln Phe Glu  
 1                      5                      10                      15  
 Val Asn Ser Val Asp Gln Ser Val Leu Ala Ser Tyr Gln Thr Leu Gln  
 20                      25                      30  
 Pro Glu Leu Asn Ala Leu Leu Asn Ser Ala Pro Thr Pro Glu Met Leu  
 35                      40                      45

Ser Ile Thr Ile Ser Asp Asp Ser Asp Ala Asn Ser Phe Glu Ser Gln  
 50 55 60

Leu Asn Ala Ala Thr Asn Ala Ile Asn Asn Gly Tyr Ile Val Lys Leu  
 65 70 75 80

Ala Thr Ala Thr His Ala Leu Leu Met Leu Pro Ala Leu Lys Ala Ala  
 85 90 95

Gln Met Arg Ile His Pro His Ala Gln Leu Ala Ala Met Gln Gln Ala  
 100 105 110

Lys Ser Thr Pro Met Ser Gln Val Ser Gly Glu Leu Lys Leu Gly Ala  
 115 120 125

Asn Ala Leu Ser Leu Ala Gln Thr Asn Ala Leu Ser His Ala Leu Ser  
 130 135 140

Gln Ala Lys Arg Asn Leu Thr Asp Val Ser Val Asn Glu Cys Phe Glu  
 145 150 155 160

Asn Leu Lys Ser Glu Gln Gln Phe Thr Glu Val Tyr Ser Leu Ile Gln  
 165 170 175

Gln Leu Ala Ser Arg Thr His Val Arg Lys Glu Val Asn Gln Gly Val  
 180 185 190

Glu Leu Gly Pro Lys Gln Ala Lys Ser His Tyr Trp Phe Ser Glu Phe  
 195 200 205

His Gln Asn Arg Val Ala Ala Ile Asn Phe Ile Asn Gly Gln Gln Ala  
 210 215 220

Thr Ser Tyr Val Leu Thr Gln Gly Ser Gly Leu Leu Ala Ala Lys Ser  
 225 230 235 240

Met Leu Asn Gln Gln Arg Leu Met Phe Ile Leu Pro Gly Asn Ser Gln  
 245 250 255

Gln Gln Ile Thr Ala Ser Ile Thr Gln Leu Met Gln Gln Leu Glu Arg  
 260 265 270

Leu Gln Val Thr Glu Val Asn Glu Leu Ser Leu Glu Cys Gln Leu Glu  
 275 280 285

Leu Leu Ser Ile Met Tyr Asp Asn Leu Val Asn Ala Asp Lys Leu Thr  
 290 295 300

Thr Arg Asp Ser Lys Pro Ala Tyr Gln Ala Val Ile Gln Ala Ser Ser  
 305 310 315 320  
  
 Val Ser Ala Ala Lys Gln Glu Leu Ser Ala Leu Asn Asp Ala Leu Thr  
 325 330 335  
  
 Ala Leu Phe Ala Glu Gln Thr Asn Ala Thr Ser Thr Asn Lys Gly Leu  
 340 345 350  
  
 Ile Gln Tyr Lys Thr Pro Ala Gly Ser Tyr Leu Thr Leu Thr Pro Leu  
 355 360 365  
  
 Gly Ser Asn Asn Asp Asn Ala Gln Ala Gly Leu Ala Phe Val Tyr Pro  
 370 375 380  
  
 Gly Val Gly Thr Val Tyr Ala Asp Met Leu Asn Glu Leu His Gln Tyr  
 385 390 395 400  
  
 Phe Pro Ala Leu Tyr Ala Lys Leu Glu Arg Glu Gly Asp Leu Lys Ala  
 405 410 415  
  
 Met Leu Gln Ala Glu Asp Ile Tyr His Leu Asp Pro Lys His Ala Ala  
 420 425 430  
  
 Gln Met Ser Leu Gly Asp Leu Ala Ile Ala Gly Val Gly Ser Ser Tyr  
 435 440 445  
  
 Leu Leu Thr Gln Leu Leu Thr Asp Glu Phe Asn Ile Lys Pro Asn Phe  
 450 455 460  
  
 Ala Leu Gly Tyr Ser Met Gly Glu Ala Ser Met Trp Ala Ser Leu Gly  
 465 470 475 480  
  
 Val Trp Gln Asn Pro His Ala Leu Ile Ser Lys Thr Gln Thr Asp Pro  
 485 490 495  
  
 Leu Phe Thr Ser Ala Ile Ser Gly Lys Leu Thr Ala Val Arg Gln Ala  
 500 505 510  
  
 Trp Gln Leu Asp Asp Thr Ala Ala Glu Ile Gln Trp Asn Ser Phe Val  
 515 520 525  
  
 Val Arg Ser Glu Ala Ala Pro Ile Glu Ala Leu Leu Lys Asp Tyr Pro  
 530 535 540  
  
 His Ala Tyr Leu Ala Ile Ile Gln Gly Asp Thr Cys Val Ile Ala Gly  
 545 550 555 560

Cys Glu Ile Gln Cys Lys Ala Leu Leu Ala Ala Leu Gly Lys Arg Gly  
                             565                            570                            575

Ile Ala Ala Asn Arg Val Thr Ala Met His Thr Gln Pro Ala Met Gln  
                             580                            585                            590

Glu His Gln Asn Val Met Asp Phe Tyr Leu Gln Pro Leu Lys Ala Glu  
                             595                            600                            605

Leu Pro Ser Glu Ile Ser Phe Ile Ser Ala Ala Asp Leu Thr Ala Lys  
                             610                            615                            620

Gln Thr Val Ser Glu Gln Ala Leu Ser Ser Gln Val Val Ala Gln Ser  
                             625                            630                            635                            640

Ile Ala Asp Thr Phe Cys Gln Thr Leu Asp Phe Thr Ala Leu Val His  
                             645                            650                            655

His Ala Gln His Gln Gly Ala Lys Leu Phe Val Glu Ile Gly Ala Asp  
                             660                            665                            670

Arg Gln Asn Cys Thr Leu Ile Asp Lys Ile Val Lys Gln Asp Gly Ala  
                             675                            680                            685

Ser Ser Val Gln His Gln Pro Cys Cys Thr Val Pro Met Asn Ala Lys  
                             690                            695                            700

Gly Ser Gln Asp Ile Thr Ser Val Ile Lys Ala Leu Gly Gln Leu Ile  
                             705                            710                            715                            720

Ser His Gln Val Pro Leu Ser Val Gln Pro Phe Ile Asp Gly Leu Lys  
                             725                            730                            735

Arg Glu Leu Thr Leu Cys Gln Leu Thr Ser Gln Gln Leu Ala Ala His  
                             740                            745                            750

Ala Asn Val Asp Ser Lys Phe Glu Ser Asn Gln Asp His Leu Leu Gln  
                             755                            760                            765

Gly Glu Val  
                             770

&lt;210&gt; 9

&lt;211&gt; 2004

&lt;212&gt; PRT

&lt;213&gt; Shewanella putrefaciens

&lt;400&gt; 9

Met Ser Leu Pro Asp Asn Ala Ser Asn His Leu Ser Ala Asn Gln Lys  
 1 5 10 15

Gly Ala Ser Gln Ala Ser Lys Thr Ser Lys Gln Ser Lys Ile Ala Ile  
 20 25 30

Val Gly Leu Ala Thr Leu Tyr Pro Asp Ala Lys Thr Pro Gln Glu Phe  
 35 40 45

Trp Gln Asn Leu Leu Asp Lys Arg Asp Ser Arg Ser Thr Leu Thr Asn  
 50 55 60

Glu Lys Leu Gly Ala Asn Ser Gln Asp Tyr Gln Gly Val Gln Gly Gln  
 65 70 75 80

Ser Asp Arg Phe Tyr Cys Asn Lys Gly Gly Tyr Ile Glu Asn Phe Ser  
 85 90 95

Phe Asn Ala Ala Gly Tyr Lys Leu Pro Glu Gln Ser Leu Asn Gly Leu  
 100 105 110

Asp Asp Ser Phe Leu Trp Ala Leu Asp Thr Ser Arg Asn Ala Leu Ile  
 115 120 125

Asp Ala Gly Ile Asp Ile Asn Gly Ala Asp Leu Ser Arg Ala Gly Val  
 130 135 140

Val Met Gly Ala Leu Ser Phe Pro Thr Thr Arg Ser Asn Asp Leu Phe  
 145 150 155 160

Leu Pro Ile Tyr His Ser Ala Val Glu Lys Ala Leu Gln Asp Lys Leu  
 165 170 175

Gly Val Lys Ala Phe Lys Leu Ser Pro Thr Asn Ala His Thr Ala Arg  
 180 185 190

Ala Ala Asn Glu Ser Ser Leu Asn Ala Ala Asn Gly Ala Ile Ala His  
 195 200 205

Asn Ser Ser Lys Val Val Ala Asp Ala Leu Gly Leu Gly Gly Ala Gln  
 210 215 220

Leu Ser Leu Asp Ala Ala Cys Ala Ser Ser Val Tyr Ser Leu Lys Leu  
 225 230 235 240

Ala Cys Asp Tyr Leu Ser Thr Gly Lys Ala Asp Ile Met Leu Ala Gly  
 245 250 255

Ala Val Ser Gly Ala Asp Pro Phe Phe Ile Asn Met Gly Phe Ser Ile  
 260 265 270

Phe His Ala Tyr Pro Asp His Gly Ile Ser Val Pro Phe Asp Ala Ser  
 275 280 285

Ser Lys Gly Leu Phe Ala Gly Glu Gly Ala Gly Val Leu Val Leu Lys  
 290 295 300

Arg Leu Glu Asp Ala Glu Arg Asp Asn Asp Lys Ile Tyr Ala Val Val  
 305 310 315 320

Ser Gly Val Gly Leu Ser Asn Asp Gly Lys Gly Gln Phe Val Leu Ser  
 325 330 335

Pro Asn Pro Lys Gly Gln Val Lys Ala Phe Glu Arg Ala Tyr Ala Ala  
 340 345 350

Ser Asp Ile Glu Pro Lys Asp Ile Glu Val Ile Glu Cys His Ala Thr  
 355 360 365

Gly Thr Pro Leu Gly Asp Lys Ile Glu Leu Thr Ser Met Glu Thr Phe  
 370 375 380

Phe Glu Asp Lys Leu Gln Gly Thr Asp Ala Pro Leu Ile Gly Ser Ala  
 385 390 395 400

Lys Ser Asn Leu Gly His Leu Leu Thr Ala Ala His Ala Gly Ile Met  
 405 410 415

Lys Met Ile Phe Ala Met Lys Glu Gly Tyr Leu Pro Pro Ser Ile Asn  
 420 425 430

Ile Ser Asp Ala Ile Ala Ser Pro Lys Lys Leu Phe Gly Lys Pro Thr  
 435 440 445

Leu Pro Ser Met Val Gln Gly Trp Pro Asp Lys Pro Ser Asn Asn His  
 450 455 460

Phe Gly Val Arg Thr Arg His Ala Gly Val Ser Val Phe Gly Phe Gly  
 465 470 475 480

Gly Cys Asn Ala His Leu Leu Leu Glu Ser Tyr Asn Gly Lys Gly Thr  
 485 490 495

Val Lys Ala Glu Ala Thr Gln Val Pro Arg Gln Ala Glu Pro Leu Lys  
 500 505 510

Val Val Gly Leu Ala Ser His Phe Gly Pro Leu Ser Ser Ile Asn Ala  
 515 520 525  
 Leu Asn Asn Ala Val Thr Gln Asp Gly Asn Gly Phe Ile Glu Leu Pro  
 530 535 540  
 Lys Lys Arg Trp Lys Gly Leu Glu Lys His Ser Glu Leu Leu Ala Glu  
 545 550 555 560  
 Phe Gly Leu Ala Ser Ala Pro Lys Gly Ala Tyr Val Asp Asn Phe Glu  
 565 570 575  
 Leu Asp Phe Leu Arg Phe Lys Leu Pro Pro Asn Glu Asp Asp Arg Leu  
 580 585 590  
 Ile Ser Gln Gln Leu Met Leu Met Arg Val Thr Asp Glu Ala Ile Arg  
 595 600 605  
 Asp Ala Lys Leu Glu Pro Gly Gln Lys Val Ala Val Leu Val Ala Met  
 610 615 620  
 Glu Thr Glu Leu Glu Leu His Gln Phe Arg Gly Arg Val Asn Leu His  
 625 630 635 640  
 Thr Gln Leu Ala Gln Ser Leu Ala Ala Met Gly Val Ser Leu Ser Thr  
 645 650 655  
 Asp Glu Tyr Gln Ala Leu Glu Ala Ile Ala Met Asp Ser Val Leu Asp  
 660 665 670  
 Ala Ala Lys Leu Asn Gln Tyr Thr Ser Phe Ile Gly Asn Ile Met Ala  
 675 680 685  
 Ser Arg Val Ala Ser Leu Trp Asp Phe Asn Gly Pro Ala Phe Thr Ile  
 690 695 700  
 Ser Ala Ala Glu Gln Ser Val Ser Arg Cys Ile Asp Val Ala Gln Asn  
 705 710 715 720  
 Leu Ile Met Glu Asp Asn Leu Asp Ala Val Val Ile Ala Ala Val Asp  
 725 730 735  
 Leu Ser Gly Ser Phe Glu Gln Val Ile Leu Lys Asn Ala Ile Ala Pro  
 740 745 750  
 Val Ala Ile Glu Pro Asn Leu Glu Ala Ser Leu Asn Pro Thr Ser Ala  
 755 760 765



Ser Trp Asn Val Gly Glu Gly Ala Gly Ala Val Val Leu Val Lys Asn  
 770 775 780

Glu Ala Thr Ser Gly Cys Ser Tyr Gly Gln Ile Asp Ala Leu Gly Phe  
 785 790 795 800

Ala Lys Thr Ala Glu Thr Ala Leu Ala Thr Asp Lys Leu Leu Ser Gln  
 805 810 815

Thr Ala Thr Asp Phe Asn Lys Val Lys Val Ile Glu Thr Met Ala Ala  
 820 825 830

Pro Ala Ser Gln Ile Gln Leu Ala Pro Ile Val Ser Ser Gln Val Thr  
 835 840 845

His Thr Ala Ala Glu Gln Arg Val Gly His Cys Phe Ala Ala Ala Gly  
 850 855 860

Met Ala Ser Leu Leu His Gly Leu Leu Asn Leu Asn Thr Val Ala Gln  
 865 870 875 880

Thr Asn Lys Ala Asn Cys Ala Leu Ile Asn Asn Ile Ser Glu Asn Gln  
 885 890 895

Leu Ser Gln Leu Leu Ile Ser Gln Thr Ala Ser Glu Gln Gln Ala Leu  
 900 905 910

Thr Ala Arg Leu Ser Asn Glu Leu Lys Ser Asp Ala Lys His Gln Leu  
 915 920 925

Val Lys Gln Val Thr Leu Gly Gly Arg Asp Ile Tyr Gln His Ile Val  
 930 935 940

Asp Thr Pro Leu Ala Ser Leu Glu Ser Ile Thr Gln Lys Leu Ala Gln  
 945 950 955 960

Ala Thr Ala Ser Thr Val Val Asn Gln Val Lys Pro Ile Lys Ala Ala  
 965 970 975

Gly Ser Val Glu Met Ala Asn Ser Phe Glu Thr Glu Ser Ser Ala Glu  
 980 985 990

Pro Gln Ile Thr Ile Ala Ala Gln Gln Thr Ala Asn Ile Gly Val Thr  
 995 1000 1005

Ala Gln Ala Thr Lys Arg Glu Leu Gly Thr Pro Pro Met Thr Thr Asn  
 1010 1015 1020

Thr Ile Ala Asn Thr Ala Asn Asn Leu Asp Lys Thr Leu Glu Thr Val  
 1025 1030 1035 1040  
 Ala Gly Asn Thr Val Ala Ser Lys Val Gly Ser Gly Asp Ile Val Asn  
 1045 1050 1055  
 Phe Gln Gln Asn Gln Gln Leu Ala Gln Gln Ala His Leu Ala Phe Leu  
 1060 1065 1070  
 Glu Ser Arg Ser Ala Gly Met Lys Val Ala Asp Ala Leu Leu Lys Gln  
 1075 1080 1085  
 Gln Leu Ala Gln Val Thr Gly Gln Thr Ile Asp Asn Gln Ala Leu Asp  
 1090 1095 1100  
 Thr Gln Ala Val Asp Thr Gln Thr Ser Glu Asn Val Ala Ile Ala Ala  
 1105 1110 1115 1120  
 Glu Ser Pro Val Gln Val Thr Thr Pro Val Gln Val Thr Thr Pro Val  
 1125 1130 1135  
 Gln Ile Ser Val Val Glu Leu Lys Pro Asp His Ala Asn Val Pro Pro  
 1140 1145 1150  
 Tyr Thr Pro Pro Val Pro Ala Leu Lys Pro Cys Ile Trp Asn Tyr Ala  
 1155 1160 1165  
 Asp Leu Val Glu Tyr Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Ser  
 1170 1175 1180  
 Asp Tyr Ala Ile Ile Asp Ser Tyr Ser Arg Arg Val Arg Leu Pro Thr  
 1185 1190 1195 1200  
 Thr Asp Tyr Leu Leu Val Ser Arg Val Thr Lys Leu Asp Ala Thr Ile  
 1205 1210 1215  
 Asn Gln Phe Lys Pro Cys Ser Met Thr Thr Glu Tyr Asp Ile Pro Val  
 1220 1225 1230  
 Asp Ala Pro Tyr Leu Val Asp Gly Gln Ile Pro Trp Ala Val Ala Val  
 1235 1240 1245  
 Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Leu Gly Ile Asp  
 1250 1255 1260  
 Phe Glu Asn Lys Gly Glu Arg Val Tyr Arg Leu Leu Asp Cys Thr Leu  
 1265 1270 1275 1280

Thr Phe Leu Gly Asp Leu Pro Arg Gly Gly Asp Thr Leu Arg Tyr Asp  
 1285 1290 1295

Ile Lys Ile Asn Asn Tyr Ala Arg Asn Gly Asp Thr Leu Leu Phe Phe  
 1300 1305 1310

Phe Ser Tyr Glu Cys Phe Val Gly Asp Lys Met Ile Leu Lys Met Asp  
 1315 1320 1325

Gly Gly Cys Ala Gly Phe Phe Thr Asp Glu Glu Leu Ala Asp Gly Lys  
 1330 1335 1340

Gly Val Ile Arg Thr Glu Glu Glu Ile Lys Ala Arg Ser Leu Val Gln  
 1345 1350 1355 1360

Lys Gln Arg Phe Asn Pro Leu Leu Asp Cys Pro Lys Thr Gln Phe Ser  
 1365 1370 1375

Tyr Gly Asp Ile His Lys Leu Leu Thr Ala Asp Ile Glu Gly Cys Phe  
 1380 1385 1390

Gly Pro Ser His Ser Gly Val His Gln Pro Ser Leu Cys Phe Ala Ser  
 1395 1400 1405

Glu Lys Phe Leu Met Ile Glu Gln Val Ser Lys Val Asp Arg Thr Gly  
 1410 1415 1420

Gly Thr Trp Gly Leu Gly Leu Ile Glu Gly His Lys Gln Leu Glu Ala  
 1425 1430 1435 1440

Asp His Trp Tyr Phe Pro Cys His Phe Lys Gly Asp Gln Val Met Ala  
 1445 1450 1455

Gly Ser Leu Met Ala Glu Gly Cys Gly Gln Leu Leu Gln Phe Tyr Met  
 1460 1465 1470

Leu His Leu Gly Met His Thr Gln Thr Lys Asn Gly Arg Phe Gln Pro  
 1475 1480 1485

Leu Glu Asn Ala Ser Gln Gln Val Arg Cys Arg Gly Gln Val Leu Pro  
 1490 1495 1500

Gln Ser Gly Val Leu Thr Tyr Arg Met Glu Val Thr Glu Ile Gly Phe  
 1505 1510 1515 1520

Ser Pro Arg Pro Tyr Ala Lys Ala Asn Ile Asp Ile Leu Leu Asn Gly  
 1525 1530 1535

Lys Ala Val Val Asp Phe Gln Asn Leu Gly Val Met Ile Lys Glu Glu  
 1540 1545 1550  
 Asp Glu Cys Thr Arg Tyr Pro Leu Leu Thr Glu Ser Thr Thr Ala Ser  
 1555 1560 1565  
 Thr Ala Gln Val Asn Ala Gln Thr Ser Ala Lys Lys Val Tyr Lys Pro  
 1570 1575 1580  
 Ala Ser Val Asn Ala Pro Leu Met Ala Gln Ile Pro Asp Leu Thr Lys  
 1585 1590 1595 1600  
 Glu Pro Asn Lys Gly Val Ile Pro Ile Ser His Val Glu Ala Pro Ile  
 1605 1610 1615  
 Thr Pro Asp Tyr Pro Asn Arg Val Pro Asp Thr Val Pro Phe Thr Pro  
 1620 1625 1630  
 Tyr His Met Phe Glu Phe Ala Thr Gly Asn Ile Glu Asn Cys Phe Gly  
 1635 1640 1645  
 Pro Glu Phe Ser Ile Tyr Arg Gly Met Ile Pro Pro Arg Thr Pro Cys  
 1650 1655 1660  
 Gly Asp Leu Gln Val Thr Thr Arg Val Ile Glu Val Asn Gly Lys Arg  
 1665 1670 1675 1680  
 Gly Asp Phe Lys Lys Pro Ser Ser Cys Ile Ala Glu Tyr Glu Val Pro  
 1685 1690 1695  
 Ala Asp Ala Trp Tyr Phe Asp Lys Asn Ser His Gly Ala Val Met Pro  
 1700 1705 1710  
 Tyr Ser Ile Leu Met Glu Ile Ser Leu Gln Pro Asn Gly Phe Ile Ser  
 1715 1720 1725  
 Gly Tyr Met Gly Thr Thr Leu Gly Phe Pro Gly Leu Glu Leu Phe Phe  
 1730 1735 1740  
 Arg Asn Leu Asp Gly Ser Gly Glu Leu Leu Arg Glu Val Asp Leu Arg  
 1745 1750 1755 1760  
 Gly Lys Thr Ile Arg Asn Asp Ser Arg Leu Leu Ser Thr Val Met Ala  
 1765 1770 1775  
 Gly Thr Asn Ile Ile Gln Ser Phe Ser Phe Glu Leu Ser Thr Asp Gly  
 1780 1785 1790

Glu Pro Phe Tyr Arg Gly Thr Ala Val Phe Gly Tyr Phe Lys Gly Asp  
 1795 1800 1805

Ala Leu Lys Asp Gln Leu Gly Leu Asp Asn Gly Lys Val Thr Gln Pro  
 1810 1815 1820

Trp His Val Ala Asn Gly Val Ala Ala Ser Thr Lys Val Asn Leu Leu  
 1825 1830 1835 1840

Asp Lys Ser Cys Arg His Phe Asn Ala Pro Ala Asn Gln Pro His Tyr  
 1845 1850 1855

Arg Leu Ala Gly Gly Gln Leu Asn Phe Ile Asp Ser Val Glu Ile Val  
 1860 1865 1870

Asp Asn Gly Gly Thr Glu Gly Leu Gly Tyr Leu Tyr Ala Glu Arg Thr  
 1875 1880 1885

Ile Asp Pro Ser Asp Trp Phe Phe Gln Phe His Phe His Gln Asp Pro  
 1890 1895 1900

Val Met Pro Gly Ser Leu Gly Val Glu Ala Ile Ile Glu Thr Met Gln  
 1905 1910 1915 1920

Ala Tyr Ala Ile Ser Lys Asp Leu Gly Ala Asp Phe Lys Asn Pro Lys  
 1925 1930 1935

Phe Gly Gln Ile Leu Ser Asn Ile Lys Trp Lys Tyr Arg Gly Gln Ile  
 1940 1945 1950

Asn Pro Leu Asn Lys Gln Met Ser Met Asp Val Ser Ile Thr Ser Ile  
 1955 1960 1965

Lys Asp Glu Asp Gly Lys Lys Val Ile Thr Gly Asn Ala Ser Leu Ser  
 1970 1975 1980

Lys Asp Gly Leu Arg Ile Tyr Glu Val Phe Asp Ile Ala Ile Ser Ile  
 1985 1990 1995 2000

Glu Glu Ser Val

<210> 10

<211> 543

<212> PRT

<213> *Shewanella putrefaciens*

&lt;400&gt; 10

Met Asn Pro Thr Ala Thr Asn Glu Met Leu Ser Pro Trp Pro Trp Ala  
 1 5 10 15

Val Thr Glu Ser Asn Ile Ser Phe Asp Val Gln Val Met Glu Gln Gln  
 20 25 30

Leu Lys Asp Phe Ser Arg Ala Cys Tyr Val Val Asn His Ala Asp His  
 35 40 45

Gly Phe Gly Ile Ala Gln Thr Ala Asp Ile Val Thr Glu Gln Ala Ala  
 50 55 60

Asn Ser Thr Asp Leu Pro Val Ser Ala Phe Thr Pro Ala Leu Gly Thr  
 65 70 75 80

Glu Ser Leu Gly Asp Asn Asn Phe Arg Arg Val His Gly Val Lys Tyr  
 85 90 95

Ala Tyr Tyr Ala Gly Ala Met Ala Asn Gly Ile Ser Ser Glu Glu Leu  
 100 105 110

Val Ile Ala Leu Gly Gln Ala Gly Ile Leu Cys Gly Ser Phe Gly Ala  
 115 120 125

Ala Gly Leu Ile Pro Ser Arg Val Glu Ala Ala Ile Asn Arg Ile Gln  
 130 135 140

Ala Ala Leu Pro Asn Gly Pro Tyr Met Phe Asn Leu Ile His Ser Pro  
 145 150 155 160

Ser Glu Pro Ala Leu Glu Arg Gly Ser Val Glu Leu Phe Leu Lys His  
 165 170 175

Lys Val Arg Thr Val Glu Ala Ser Ala Phe Leu Gly Leu Thr Pro Gln  
 180 185 190

Ile Val Tyr Tyr Arg Ala Ala Gly Leu Ser Arg Asp Ala Gln Gly Lys  
 195 200 205

Val Val Val Gly Asn Lys Val Ile Ala Lys Val Ser Arg Thr Glu Val  
 210 215 220

Ala Glu Lys Phe Met Met Pro Ala Pro Ala Lys Met Leu Gln Lys Leu  
 225 230 235 240

Val Asp Asp Gly Ser Ile Thr Ala Glu Gln Met Glu Leu Ala Gln Leu

245								250				255			
Val	Pro	Met	Ala	Asp	Asp	Ile	Thr	Ala	Glu	Ala	Asp	Ser	Gly	Gly	His
260								265				270			
Thr	Asp	Asn	Arg	Pro	Leu	Val	Thr	Leu	Leu	Pro	Thr	Ile	Leu	Ala	Leu
275								280				285			
Lys	Glu	Glu	Ile	Gln	Ala	Lys	Tyr	Gln	Tyr	Asp	Thr	Pro	Ile	Arg	Val
290								295				300			
Gly	Cys	Gly	Gly	Gly	Val	Gly	Thr	Pro	Asp	Ala	Ala	Leu	Ala	Thr	Phe
305								310				315			
Asn	Met	Gly	Ala	Ala	Tyr	Ile	Val	Thr	Gly	Ser	Ile	Asn	Gln	Ala	Cys
325								330				335			
Val	Glu	Ala	Gly	Ala	Ser	Asp	His	Thr	Arg	Lys	Leu	Leu	Ala	Thr	Thr
340								345				350			
Glu	Met	Ala	Asp	Val	Thr	Met	Ala	Pro	Ala	Ala	Asp	Met	Phe	Glu	Met
355								360				365			
Gly	Val	Lys	Leu	Gln	Val	Val	Lys	Arg	Gly	Thr	Leu	Phe	Pro	Met	Arg
370								375				380			
Ala	Asn	Lys	Leu	Tyr	Glu	Ile	Tyr	Thr	Arg	Tyr	Asp	Ser	Ile	Glu	Ala
385								390				395			
Ile	Pro	Leu	Asp	Glu	Arg	Glu	Lys	Leu	Glu	Lys	Gln	Val	Phe	Arg	Ser
405								410				415			
Ser	Leu	Asp	Glu	Ile	Trp	Ala	Gly	Thr	Val	Ala	His	Phe	Asn	Glu	Arg
420								425				430			
Asp	Pro	Lys	Gln	Ile	Glu	Arg	Ala	Glu	Gly	Asn	Pro	Lys	Arg	Lys	Met
435								440				445			
Ala	Leu	Ile	Phe	Arg	Trp	Tyr	Leu	Gly	Leu	Ser	Ser	Arg	Trp	Ser	Asn
450								455				460			
Ser	Gly	Glu	Val	Gly	Arg	Glu	Met	Asp	Tyr	Gln	Ile	Trp	Ala	Gly	Pro
465								470				475			
Ala	L	u	Gly	Ala	Phe	Asn	Gln	Trp	Ala	Lys	Gly	Ser	Tyr	Leu	Asp
485								490				495			
Tyr	Gln	Asp	Arg	Asn	Ala	Val	Asp	Leu	Ala	Lys	His	Leu	Met	Tyr	Gly

500	505	510
Ala Ala Tyr Leu Asn Arg Ile Asn Ser Leu Thr Ala Gln Gly Val Lys 515	520	525
Val Pro Ala Gln Leu Leu Arg Trp Lys Pro Asn Gln Arg Met Ala 530	535	540
 <210> 11		
<211> 499		
<212> PRT		
<213> Shewanella putrefaciens		
 <400> 11		
Met Arg Lys Pro Leu Gln Thr Ile Asn Tyr Asp Tyr Ala Val Trp Asp 1	5	10 15
Arg Thr Tyr Ser Tyr Met Lys Ser Asn Ser Ala Ser Ala Lys Arg Tyr 20	25	30
Tyr Glu Lys His Glu Tyr Pro Asp Asp Thr Phe Lys Ser Leu Lys Val 35	40	45
Asp Gly Val Phe Ile Phe Asn Arg Thr Asn Gln Pro Val Phe Ser Lys 50	55	60
Gly Phe Asn His Arg Asn Asp Ile Pro Leu Val Phe Glu Leu Thr Asp 65	70	75 80
Phe Lys Gln His Pro Gln Asn Ile Ala Leu Ser Pro Gln Thr Lys Gln 85	90	95
Ala His Pro Pro Ala Ser Lys Pro Leu Asp Ser Pro Asp Asp Val Pro 100	105	110
Ser Thr His Gly Val Ile Ala Thr Arg Tyr Gly Pro Ala Ile Tyr Tyr 115	120	125
Ser Ser Thr Ser Ile Leu Lys Ser Asp Arg Ser Gly Ser Gln Leu Gly 130	135	140
Tyr Leu Val Phe Ile Arg Leu Ile Asp Glu Trp Phe Ile Ala Glu Leu 145	150	155 160
Ser Gln Tyr Thr Ala Ala Gly Val Glu Ile Ala Met Ala Asp Ala Ala 165	170	175



Asp Ala Gln Leu Ala Arg Leu Gly Ala Asn Thr Lys Leu Asn Lys Val  
 180 185 190

Thr Ala Thr Ser Glu Arg Leu Ile Thr Asn Val Asp Gly Lys Pro Leu  
 195 200 205

Leu Lys Leu Val Leu Tyr His Thr Asn Asn Gln Pro Pro Pro Met Leu  
 210 215 220

Asp Tyr Ser Ile Ile Ile Leu Leu Val Glu Met Ser Phe Leu Leu Ile  
 225 230 235 240

Leu Ala Tyr Phe Leu Tyr Ser Tyr Phe Leu Val Arg Pro Val Arg Lys  
 245 250 255

Leu Ala Ser Asp Ile Lys Lys Met Asp Lys Ser Arg Glu Ile Lys Lys  
 260 265 270

Leu Arg Tyr His Tyr Pro Ile Thr Glu Leu Val Lys Val Ala Thr His  
 275 280 285

Phe Asn Ala Leu Met Gly Thr Ile Gln Glu Gln Thr Lys Gln Leu Asn  
 290 295 300

Glu Gln Val Phe Ile Asp Lys Leu Thr Asn Ile Pro Asn Arg Arg Ala  
 305 310 315 320

Phe Glu Gln Arg Leu Glu Thr Tyr Cys Gln Leu Leu Ala Arg Gln Gln  
 325 330 335

Ile Gly Phe Thr Leu Ile Ile Ala Asp Val Asp His Phe Lys Glu Tyr  
 340 345 350

Asn Asp Thr Leu Gly His Leu Ala Gly Asp Glu Ala Leu Ile Lys Val  
 355 360 365

Ala Gln Thr Leu Ser Gln Gln Phe Tyr Arg Ala Glu Asp Ile Cys Ala  
 370 375 380

Arg Phe Gly Gly Glu Glu Phe Ile Met Leu Phe Arg Asp Ile Pro Asp  
 385 390 395 400

Glu Pro Leu Gln Arg Lys Leu Asp Ala Met Leu His Ser Phe Ala Glu  
 405 410 415

Leu Asn Leu Pro His Pro Asn Ser Ser Thr Ala Asn Tyr Val Thr Val  
 420 425 430

Ser Leu Gly Val Cys Thr Val Val Ala Val Asp Asp Phe Glu Phe Lys  
 435 440 445

Ser Glu Ser His Ile Ile Gly Ser Gln Ala Ala Leu Ile Ala Asp Lys  
 450 455 460

Ala Leu Tyr His Ala Lys Ala Cys Gly Arg Asn Gln Ala Leu Ser Lys  
 465 470 475 480

Thr Thr Ile Thr Val Asp Glu Ile Glu Gln Leu Glu Ala Asn Lys Ile  
 485 490 495

Gly His Gln

<210> 12

<211> 40138

<212> DNA

<213> *Vibrio marinus*

<400>.12

aatagatcga ctcgcaaaag ttgcttaaga tagtgtcaat atagcttctt atttgtaaatt 60  
 attgtttttt atgtgtaaac atgttttagt tgtgtaaatg ctgttaatta tccttttggg 120  
 attgtaatag ctgatgttgc tggctaataa gtacttttag ttcggaataa tcttgcttta 180  
 aatcgctaac ttcagttttt aattcaccca cacttggtgt atttttaagg ctctcttccc 240  
 caccatcgac aaaccaggat gatatgaaac cgttaaactg accaaagaga ccgacacctg 300  
 cagtcattgag taatgccgca atgatacgtc cgcagtggtg gacggggtag tagtcacctg 360  
 aaccaacagt cgttattgtc acaaatgacc accaaagtgc gtcgatgccg ttattgatgt 420  
 tactgcctac ttgatcctgt tctaacaata aaataccgat agcaccaaag gtgacaagga 480  
 tgaaggatat cgcagatacc agcgaagagg tggctttaaa ccgatgttca aaaatcattt 540  
 ttaagataat ttttgatgag cgtatattct gaatagatct taatactcta gcgatacga 600  
 ttatgcgaat aaactgcagt tgctcgacca tcggaatact cgacagttag tcaatccaac 660  
 cccatttcat aaactgaaat ttattctcag cttggtgaaa gcgaattaca aagtcagtga 720  
 aaaagaataa gcaaatacgt ttatctacgc tcgttaatat ttcagtgcag ttacttgaaa 780  
 aggtaaaaat aagttgcagt agtgatgata cgaccacatg aagtgaataa ataagcatga 840  
 aaatctgaaa tggatttaca tcactgttgt ttttggtgcc acttttaagg ttcgctttca 900  
 caatctgctg cctcgggttca ttgattttgt taatataaac cttagtcagt agcaagacaa 960  
 aatataattta catcaatgtc atcgtattat tcaaccgcgc gtcgtgtatt cagaccaaga 1020  
 tcgttgatata tgtagtcatt gtagcgatga gattatcatg cgacaggaga gaattatggt 1080  
 tgttattatt ttttacgtac cttaaagttta tgttgaaaga gtaaaacagg cgttatttaa 1140  
 cgtcggagct ggcaccatcg gtgattatga tagttgtgct tggcaatgtt tggggactgg 1200  
 gcagttccaa cctttacttg gtagccagcc acatattggt aagctaaatg aggttgaaatt 1260  
 cgttgatgag ttttagagtag aaatgggttg tcgagcagaa aatgtaaggg cagcaataaa 1320  
 tgcacttatt gctgcgcacc cttatgaaga acctgcttat catattctgc aaacattgaa 1380  
 tcttgatgag ttaccttaag ttagatgcac tgcacttaat tggttcgtg tgctaggtta 1440  
 gcaattagca attttgacca tgttagcgat agttttggca caagtgatcg atattaaact 1500  
 atccgattca gatccattt ttactgctga attaggttc attacacttg ttctagtgg 1560

```

tttccccgac aggtgtaact ctgttacttg cgtaagggtg ataatctcta ccgcattggc 1620
aggagttaca cctgcaccag gcataaatact aattctacca tctgcttggc taactaacgt 1680
tcggattaaag gcgcagcctt ctagcgcttg agcttgttga ccagagggtta aaatacgcctc 1740
acaaccagca gtgatcaagg tctccaaggc ttgttgttga tcattacaca agtcgaaagc 1800
gcggttgaag gttacgcgga gatcacgtga tgccaccatt aagcgtttta aagctggctc 1860
gtcaatatta ccatctgctg ttaacgcgcc aataacgacc ccttggacac cgagtaactt 1920
catgaatttg atgtcggaaa ccataatatc aacttcttgt tcgctatata caaaatcacc 1980
ggcgcgaggg cgaataatgg cataaatggg gatcgttget agatcaatag acctttgtac 2040
aaaacctgcg ttggcgggtca agccacctaa tgctaattgc gagcacaact caatacgcac 2100
ggcgccagat gcttgagccg tcagcagtga ttctatatta tcgacacata cttctattgt 2160
cattgtcata tacttctctt taaaaagtgt attaaaaata ataaagccag cataagtcgt 2220
tttatacaat atgaaagggg aaaaggcgac ttagctcgcc tagatcaatt attatggcag 2280
aatactgccg tattgtgatt agaaagacag ttttttaagc tcaatagccg ttatcgctt 2340
gttatctacc atcgtgtaac ttttctggcc tgggtgcttt attaacactg tttcagtggc 2400
tggattaggg tgaaatgatt cttttttcaa atctgttttt ttgtatttga acgtacctgt 2460
aatgtcttgc tgctcacgaa gacgtacaaa tattggttgc gcatagcttg gtagtgccgc 2520
attgacatgt tgatagaatt cagacgtga aaattcatga atagggaat tcaaagtcag 2580
cgcgaccatg cctgctcgcc catcgtgatg tgggagcttg acaccataag ccacactttg 2640
ctcaatttgc acaaaatcgt taacttgagc ttctacttgc gtcgtggcga cattttcac 2700
ttccagcgg aatgtatcac ctaatctatc cacaaaggaa atatggcgat aaccttggta 2760
atgaacgaga tcgcccgtat taaaaataaca gtcaccgtct ttttaatactg acttaaatag 2820
ctttttatta ctttctgtgt categgtata accatcaaat ggtgaacgtt tagttatctt 2880
tgcttagcagt agccctgttt cteccgtttt tactttggtc attttccctt tcgcattata 2940
cacaggtttg tcatgttcaa tatcatattg tatgacggtt aaagcaagtg gagttaacccc 3000
cgctgtatgc ggttaagtcca gcgcattgga gaacacaaga ttacactcac tggcgccata 3060
gaattcatta atatgctcga tccccaaacg ttgttggaaa tgatcccaaa tttcggggcg 3120
taatccatta cctatgattt tctttatatt atgctgtttg tctttattgc taggcggtac 3180
atttaataaa taacggcaga gctcgccgat gtaagtaaac gcagtggcat tatgagcatg 3240
aacttcattc caaaagcgac ttgaactgaa tttttcagaa agtcgcaggg ttgctgcgct 3300
accaaacacg gcgcttaatg acactgtcag tgcattgtta tggatatagg ggagtgataa 3360
atacaataca tcatcagctg ttaagcgtaa tgatgccatc cccatgctg ccatggattt 3420
aaaccaacgg tgatggctca ttcttctgct ttttggcagt ccagttttcc ccgaggtaaa 3480
gatataaaac gcgcaatgct taagctgtat ttgtgctgtt gattcagggt tcaatactga 3540
atatcctgcg actagtgtag atatgttttt ataaccatca ctcatgtctg gcgtttctaa 3600
agcgggtacg taaaagacat tctgttgtaa tgcgatgac aatttgggtt caatattatt 3660
aatggcggat gtgtatagtt catctgcgat gagtaatttg gtatcgacca cgctaagact 3720
atgttcgagg attgaatccc gttgtgtcgt atttatcata caagcaatcg cgccaagctt 3780
gacaactgcg agggcaataa tgatggtttc aggcctgtta tcgagcatga tggcgacttt 3840
atcattttta ccaatgccgt attcatgaag gaaatgggca tattgatttg cttgcttatt 3900
caatgaatcg taactataac gctggctctt aaattgtatt gcgatcaagt cagagttatt 3960
gacagcttgc tgetctagta ataaaccaat agacataaaa cgttcgggct ttgcttgttg 4020
taagtgccat aagcctttga tgattggctt tggggttttt aatagattga tggtaacttt 4080
caggaattgt ttgccgggtta taacagtcac aagctaattc tttttatcaa gaagaggggt 4140
tatgacacca aataaatggg tcacgcgttg gtttaatttg gttagactaa atgtgttgtt 4200
ttgctgtgat aatgcgacgt tcaaacaaac ttgagaaggt aaaaaaatag catttttaaa 4260
ttgaacatca atactaatgt gttgaatatc aatcaagttt tctaactgtg cgagcacgcg 4320
tgcttttagca aacatgccat gtgctattgc tgttttaaac cccattagtt tcgctgggat 4380
aaaatgtaaa tggattggat ttgtgtcttt ggagatataa gcatacttat atacgtcaaa 4440

```

aggactaaat ttaaacaatg aaatcggctc gtaagcataa ttcgctggcg tatttactat 4500  
 tttctcaccg ctggaacgtt gagatcgttg gcacgttttt cgctgtttcg tttctgttaa 4560  
 gaatgtcgat gtacactccc acgcaaattg tccatctaca aacacatcaa tatgagtatc 4620  
 aatgaaacgt cctgtatccg ttatgtactc ctttaattaca cgacatgtgc tgcgtcaatat 4680  
 cgcgtttaat gctatcgggt gatgttgtt tatgctgatt cgataatgga ctagtccctaa 4740  
 tatagatate ggaaattgtg ttgatgtcat gaggttcctc aataatggaa agatccatcac 4800  
 aaatggataa gtaaccggta catagtgttg gttattaaac ccacagcatt taatatattg 4860  
 ctttaaattt cgctgatcta tttttgtcc actgatacta aattgctcag tacacacttg 4920  
 tgcgaccaa gtgttcatca gtgttttaac aattgtattg accactgctt tcacatataa 4980  
 aagcgagata atcggttgct ttgttaacag tgtgatctgg ttagcgtgca ttgaaataat 5040  
 tcatataaga gtatgtagca tttatgttaa tttttgttt tggaaagtga attggcgaat 5100  
 ccgtaatcgg tttatggcag ttccgtcaaa tacttcaggt aaactcgtta ctcataccat 5160  
 tgatagtgtt aaagtgattg actgaataaa gaatagagct aaaagtggaa aaattatgca 5220  
 agatgcgggt atgttattac gcattgctta tgaggcaatg aaagagttag aggttgatgt 5280  
 cattgaagta ctttctcgtt gtaacataag tgaagaaagta ctgaatgata aggatcttcg 5340  
 cacaccta atgcacaaa cacatttttg gcaagtatta gaagacatat cacaagatcc 5400  
 taacatcggc atttcacttg gtgagagaat gccagtgttc acggggcagg tattacagta 5460  
 tctttttctc agtagtccca ctttgggtac tggctgggaa cgcgcaacaa aatactttcg 5520  
 attaatcagt gatgcggcga gtgtttctat caagatggaa ggctgtgaag cgcgattatc 5580  
 tgtgaactta gatggttag cggaagatgc gaatcgtcat ttgaatgatt gcctagtcat 5640  
 cgggtgcattt aaattttgtt tatatgtgac agaaggcgaa tttaaagtaa gcaaaatagc 5700  
 ctctgctcat gctcggccga aagatattac tgcctatacc aatgtattta catgtccgat 5760  
 tgagtttgct gccgaagata attatattta tttcgatgct gatttactcg aacgtccctc 5820  
 ttcgcatgcg gagcctgagc tattcgctt acacgatcag cttgcaagcc gtaaaatagc 5880  
 caagttagaa ctgcaagatt tagtggataa agtacgtaag gttattgcac aacaacttga 5940  
 gtctgggtgt gtgactttag aaagtatcgc cactgaactt gacatgaac cacgtatgct 6000  
 aagagcgaag ttagctgaca ttgattataa ctttaatcaa atactcgtg atttttgttg 6060  
 cgagttatca aaaaaactgt tggcgaatac ggacgagctt attgatcaga ttgtctatct 6120  
 cactgggttt tctgaaccaa gtacttttta tgcgtccctt aagcgtggg ttaaaatgac 6180  
 gccaatgaa tatcggcgta gcaaaactgc ggttaggcat gctaataaac acgagtccta 6240  
 aaaattcgct gcttagtgca tagtgcatag tgcatagtgc tagtaagcca agtacaaagc 6300  
 gttaaagtta agtacttgag cgaaccatca gacaccactt actagattaa gcacctatta 6360  
 atgattgacc acaaattctg atcgtattgc ctgtgatccc tgcagcttga gggtgcgcaa 6420  
 aaaaagctat cgcttcagca acatcaactg gcttaccacc ttgttttaat gaattcatac 6480  
 gacgaccagc ttcacgaact gtaaatggaa tgcgtgctgt ctttttgtt tcaataaagc 6540  
 ctggtgcaac agcattaatg gtgatgtatt tgtctgcaag cggagtttgc attgcatcaa 6600  
 cataaccaat gactgcggcc ttagacgttg cataattagt ctgaccaaag ttaccgcgaa 6660  
 tcccactcat cgaagacaca caaacaatgc ggccatagtc gttgagcaga tcatcattta 6720  
 gcagtgcctc attgattctt tccattgccc acaagttaat atccatcagt acatcccaat 6780  
 ggttatccgg catacgtgct agcgttttgt cttttgttac cccggcatta tggacgatga 6840  
 tatcaagcga ctgttctcgc acaaagtcag caatgatatt tggggcgctc gcagcggtaa 6900  
 tatcagcaac aatgctgcta cttttcaagc aatgagctac tttttcaagg tcctgtttta 6960  
 atgccggaat gtctaagcaa ataacatgtg cgccatcacg ggcgagtgct tcagcaatag 7020  
 cagccccgat gccacgtgat gcaccagtga caagtgtgt ctttcttgt aatggttttg 7080  
 ccgtgttact tgtttcgtta ataacttcgt taataacttc gtttaataact tcgttaatag 7140  
 ccccattaat cgaaccgggt tttacgttaa taacctgtgc tgagatatag gctgattttg 7200  
 ctgaggttaa gaaacgt gc ggggcctcta ataattgtc actaccaggt tgtacataga 7260  
 taagttagca ggtactacca ttcttgcta tttcttggc gacactgcga caaaaccctt 7320

ctaa gatct ttgtacagtc gcgtagctta catcgtcaag atgttcacac ggatgacctt 7380  
acacgatcac tctgctgcat ggcgagagct gcttaattac aggttgaaaa aaacgatgta 7440  
atgcacttaa ttgcttgctg ttcttaatgc ctgaggcgtc gaagataata ccgttgaaagc 7500  
gatctgtttt agcgatagca ttaaggctaa taggtgtcgc gactaaagac gtttgattaa 7560  
attcaatatt aagatcggct aacgctgacg tgttattagg ataagaaatc gtgacttcag 7620  
catctttaaa tgtgttaaga atgggtttaa ttaatttgct gttgctggct gcgccgatga 7680  
gtaagttgcc agagatgaga tcgggtccct gatcgtagcg tgctaacgta accggctcgtg 7740  
gcagattaag cgctttaaat aaacctgatg tccacttgcc attagcgagt ttgctgatg 7800  
tatcgcgtcat tttctaatcc ttgttatagt gaacagtttg aatctcgaag atgtacatgt 7860  
gttaaaaaatt atctgatagc tatgacttat ctgccactac gtaataataa atagaccagt 7920  
tcattacatc gtaategat atagtataac taaatactaa gtaaattata atgataagac 7980  
tgttatcgta ctcggtacaa actctgatca gcaaataatc aaattagagt ttttatttta 8040  
aacttgatc aacaatgta cattaatgta tcttacgtct aatgtgctac gggcatattt 8100  
aagtcactaa attaaaggaa taaaccatga cagggtcaaac aataagaaga gtagcaatta 8160  
tcggcggtaa ccgtatcccg tttgcacgtt caaatacagc gtattcaaaa ctaagtaacc 8220  
aagatatgct gacggaaact atccgtggct tgggtggttaa atataacctc cgtggtgaaac 8280  
aactggggga agttgttgct ggtgcggtaa ttaagcattc tcgtgatttt aacttaacac 8340  
gtgaagccgt gctaagtga ggtcttgac ctgaaacgcc ttgttatgac attcaacaag 8400  
cttgtggtac tgggtctagct gcagctatcc aagtagcaaa caaaattgctg cttggtcaaa 8460  
tagaagcggg catctgctggg ggttctgata cgacatcaga tgcaccgatt gcagtcagt 8520  
aaggcatgct tagtctatta cttgagctta atcgagctaa aacgggtgaa caacgtttga 8580  
aagcactatc tcgtctacgt ctaaaacact ttgcgccact aacgcttga aataaagagc 8640  
cgcgtaacca aatggcgatg ggcgatcatt gtcaagtaac agcgaagag tggaaatctc 8700  
cacgtgaagc acaagatgca ttggcctgct caagtcacaa aaaattagct gcagcatatg 8760  
aagaagggtt ctttgatacg ttagtctcac ctatggccgg cttaaagaaa gataacgtat 8820  
tacgcgcaga tacaacagtt gagaaactgg cttaaattgaa accttgtttt gataaagtaa 8880  
acggcactat gacggcgggt aacagtacta accttaccga tggagcatca gctgtattac 8940  
ttgcaagtga agaattggga gcggcacata acctaccagt acaagcttat ctaacatttg 9000  
gtgaaacggc cgctatcgac ttcgttgata agaaagaagg tctgttaatt gcgcctgcat 9060  
acgcagtgcc aaaaatggtg aagcgtgctg gccttacatt acaagacttc gattactatg 9120  
aaatcacatga agcatttgct gcgcagtat tagcaacgct agcagcttgg gaagacgaaa 9180  
aattctgtaa agaaaaactg ggtctagatg ctgcgcttgg ttcaattgat atgaccaagt 9240  
taaacgtgaa agggagtagc ttagccacgg gtacccatt tgcgcgaact ggtggtcgtg 9300  
ttgtcgtac gctagcgcaa ttacttgatc agaaagggtc aggtcgtggt ttgatctcga 9360  
tttgtgctgc tgggtggtaaa ggtatcacgg caattttaga gaaataaacg cactgtttat 9420  
tatctattga ttaagctgtc ctgagatact ggatattttt aaataaaacg ccaatactgc 9480  
agagtattgg cgtttttttg taataccaat tcctatataa cgggtgcattt taaacactta 9540  
atttccggca ttggtatcat aaaaaagcag caccgaagtg ctgcttgatt gtagattaac 9600  
ctattaaaat agagaggcta gaattagtct tcgtatgctt cattatgtac gccagctgca 9660  
cgaccgatg gatcagcatt gttttgaaa ctttcacccc aagctaattg tttctacagtt 9720  
gaacaagcaa cggatttacc aaacggtagc catttcgctg ctgaatcacc tgggaagtga 9780  
tcttcaaaaga tggcacgata gtagtaacct tcttctgcat ctgggtgtgtt aattgggaac 9840  
ttaaatgctg cacttgctaa catttgatca gttaccgctt cttcaacgtg tactttaagt 9900  
tggtaaatcc aagaataacc aacaccatca gagaattgtt ctttttgacg ccatacaatt 9960  
tcttcaggta gtaaatcttc aaatgcttct cgaatgatgt ttttctcaat gcggctcgccc 10020  
gtgatcattt ttagttcagc gtttagacgc attgacgcac caacaaattc tttatctaag 10080  
aaaggaaacac gtgcttcgat gccccaagct gccatagatt tgtttgcacg taagcaatca 10140  
aacatatgta atttatttac ttacgtacc gtctcttcat ggaattcttt cgcatctggc 10200

gctttgtgga agtacaagta accaccgaac agttcatcag caccttcacc agaaagcacc 10260  
 atcttaatcc ccatggcttt aattttacgt gccattaggt acataggggt tgatgcacga 10320  
 attgttggtta catcgtaggt ttcaatgtgg taaatcacgt cgcgtaaagc gtcgataacct 10380  
 tcttgcacag taaattcaat tgaatgatgg atagtacctt agtgatctgc cactttttgt 10440  
 gcagcggcta aatctggaga accatttagg cctacagaga aagagtgtag ttgtggccac 10500  
 catgcttcgg ttttaccacc gtcttcaata cgacgttttg catactgttg ggtgatttget 10560  
 gaaataacag atgaatctaa cccgcctgat aataatagc cgtaaggtag atcacacatt 10620  
 aattgacgtt taactgcac ttccaaacct tgetttaacaa cgctttttatc accaccattt 10680  
 tgtgcaacgt tatcaaaatc ttccaatca cgttgataat aaggcgtgac tacaccatcc 10740  
 ttactceaca ggtaatgacc tgctgggaat tcttcaattt gagtacaaat tggcactagt 10800  
 gctttcattt cagaggcaac ataaaagtta ccgtgttcat catagcccgat ataaagaggg 10860  
 atgataccga tatggtcacg gccaatcagg taagcgtcct ctgttttcgtc atataaagcg 10920  
 aaagcaaaaa taccatttag atcatctaaa aattgtgtgc ctttttcttt atatagcgca 10980  
 agtatcaett cgcaatctga ttctgttttg aattcaaagt ctacgttcag cgttttcttt 11040  
 aaatctttgt ggttataaat ttaccatta acagcaagta cgtgtgtctt ttcttcatta 11100  
 tatagcggct gtgcaccatt atttacatcg acaatagcaa gacgttcag aactaaaaata 11160  
 gcattgtcac ttgtatagat acctgaccaa tctgggcccgc ggtgacgtag taactttgat 11220  
 agttctagtgt cttgttcgcy aagaggttta atgtctgatt tgatgtctag aattccgaat 11280  
 attgagcaca taactaatc cttctggggc tgcgtctgca gctaactttc taaatagtgt 11340  
 gtctaatttg ccacattgta gatttaatgc aaacattaat gataaaacat ttataaaaaa 11400  
 tghtaattcaa tgtggaatcg ataatttaat ggcttaaaag tgaagatcca ttaattgtga 11460  
 tggcgagggt atagacaaat gtagacctta atgaataaag caggcacgat tgaatccatt 11520  
 caacgcaaaag tggtaactaac tatgtttta aacgttaraa atagtgtttt aaagggtata 11580  
 agtaataaat ttaaaaaaata taataatcca catgcattaa atttattatg ataaaccgct 11640  
 atatctcaat ggcaatttgg gataagtgtt aaatatatgt aaaatgaatg agttgacctg 11700  
 ctttttttac actaagtgt gaaattaaag ctataggtctg ttgttagcat tgattaataa 11760  
 cgtactaaaa tacgacatct agtatagaaa ttaaaaaaac agttgggttt gatagcataa 11820  
 ctgcataaac taatcagctt attgtctgta atatttttgt aatttaataa ggtttaataa 11880  
 aattatatgt ctgataaata taaaccgtac gacctttcct ttaaaaagac gtttttctgt 11940  
 cctaagtttt ggctgtgtg gttcgggggtg ttgcaatat acttattagc ttttatgcca 12000  
 gtaaaagccgc gtgataaatt tgctcgattc atagcgaaga aattgttttag tctaaaaatg 12060  
 atggcaaaagc gtaaaaaggt agcaaagatc aatttatcta tgtgcttccc tgaaatggat 12120  
 gatagcgaac aagaccgtat aatcatgggc aatctagtta ctttttctga aactatctta 12180  
 agttatgcag agccaagtgc gcgtagtcgt gcttataacc gtgaccgtat gatagtgcac 12240  
 ggtggcgaga atttatttcc gctacttgaa caaggtaagg cttgtatctt attagtgcgc 12300  
 catagcttcg ctattgattt tgcaggttta cacattgctt cttatggcgc gccattttgt 12360  
 actatgttta acaattctga gaatgagttg ttcgattggc tgatgacacg tcaacgcgct 12420  
 atgtttggag gcactgttta tcaccgcaag gcagggctag gggctctagt taaatcactt 12480  
 aagagcgggt aaagctgtta ttacttacct gatgaagacc atggacctaa gcgtagtgt 12540  
 ttgpgcctt tatttgcgac tcaaaaagca actttacctg taatgggcaa gctagcagaa 12600  
 aaaaacaaatg cactcgttgt tcctgtttat gcggcatata atgaatcact aggtaaaatt 12660  
 gaaaccttta ttcgaccagc aatgcaaaac ttccaatcag aaagccaga acaagatgca 12720  
 gtgatgatga ataaagagat tgaagccttg attgaatgtg gtgttgatca atatatgtg 12780  
 acacttagat tattgagaac acgtccggac ggtaaaaaaa tctactaata aagtttaata 12840  
 aacaccataa tcttcgttga atatggtgtt taccctctg aataccctct aaattaataa 12900  
 caaaaaaagc catctacgta acatctaatt atgatttagc ctgcacttgc tttgttttta 12960  
 gtcttaagag cctaataaac ttgatctagg catagattct gtctttcttt acgtaacgcg 13020  
 atctattttt ttaaccgat agttgttata attagtttca tatgaaagag atatcgtttc 13080

agtaaaagct atttcgtttc aatagataat ttatttatag tcatattttc tgtaatgaca 13140  
atcattttct catctagact atagataaga atacgaatta agtaagaaca ttaattttac 13200  
aagaatataa aatatcccat cggagctata agaatgaaaa agactaaaat tgtttgtaca 13260  
attggtccaa aaactgaatc agtagagaaa ctaacagagc ttgttaatgc aggcattgaac 13320  
gttatgcgtt taaatttctc tcatggtaac ttgtctgaac attcagtgcg tattcaaaat 13380  
atccgtcaag taagtgaaaa cctgaataag aaaattgctg ttttactgga tactaaagg 13440  
ccagaaatcc gtacgattaa actagaaaac ggtgacgatg taatgttgac cgctggtcag 13500  
tcattcacgt ttacaacaga cattaacgtg gtaggtaata aagactgtgt tgctgtaaca 13560  
tatgctggtt ttgctaaaga ccttaatcct ggtgcaatca tccttgttga tgatgggtta 13620  
attgaaatgg aagttgttgc aacaactgac actgaagtta aatgtacagt attaaatact 13680  
ggtgcacttg gtgaaaaata aggcgttaac ttacctaaac tcagtgtagg tctacctgca 13740  
ttgtcagaaa aagataaagc tgatttagcg tttggtgtg agcaagaagt tgattttgtt 13800  
gctgcatcat ttattcgtaa ggctgatgat gtaagagaaa ttcgtgaaat cctatttaat 13860  
aatggtggcg aaacattca gattatctcg aaaattgaaa accaagaagg tgtagacaat 13920  
ttcgatgaaa tcttagctga atcagacggt atcatggttg ctcgtggcg tctcggtgtt 13980  
gagatcccg ttgaagaagt gatcatggca cagaagatga tgatcaaaaa atgtaataaa 14040  
gcaggtaaa ttgtaattac tgcaacacaa atgcttgatt caatgatcag taaccacgt 14100  
ccaacacgtg cagaagcggg cgatgttgcc aatgctgtgc ttgacggtac cgacgcggt 14160  
atgctttctg gtgaaactgc gaaaggtaaa taccagttg aagctgtgtc tatcatggca 14220  
aacatctgtg aacgtactga taactcaatg tcttcggatt taggtgcgaa cattgttgc 14280  
aaaagcatgc gcattacaga agctgtgtgt aaagggtgcg tagaaacaac agaaaaattg 14340  
tgtgtccac ttattgttgt tgcaactcgt ggcggtaaat cagcaaaatc tgttcgtaaa 14400  
tacttccga aagcaaatat tcttgctatc acaacaaatg aaaaagcagc gcaacagtta 14460  
tgcctaacta aaggcgtaag cagctgcac gttgagcaga ttgatagcac tgatgagttc 14520  
taccgtaaa gttaaagagct tgcattagca actggtttag cttaaagaagg cgatctcgtt 14580  
gttatggtat caggtgcgtt agtaccatca ggtacaacga atacggcatc tgttcaccaa 14640  
ctttaagttg ccataattgat attataaaaa agagagcgta tgctctctt tttatatct 14700  
gtagtttata tgtctgtaca aaaaaatgat aaagagtaca taaactatta atatagcgta 14760  
atatataatg attaacggtg atgaaaggtg taaataaatg gatagtgtca aacataaaat 14820  
tggttagtc ctttctggcg gtggtgcgaa aggtattgct catcttggtg tattaataa 14880  
cctgttagag caagatataa gaccgaatgt aattgcgggt acaagtgtc gctctatggt 14940  
tggtgcactt tattgtcag gacttgagat tgatgacatt ttacaattct tcatcgatgt 15000  
aaaacctttt tcttggaagt ttaccctgc ccgtgctggc tttatagacc cggcaaaatt 15060  
atatcctgaa gtgctaaaat atatcccgga ggatagcttt gactaccttc aacctgaatt 15120  
gcgcattgtt gccaccaaca tgttactcgg taaagagcat atatttaaag atggctccgt 15180  
gattaatgcc ttattagcat cagccagcta cccttagtt ttttctccga tgatcattga 15240  
cgatcaagtg tattcagatg gcggtattgt taatcatttc cccgtgagtg tcattgaaga 15300  
tgattgcgat aaaataatcg gcgtatacgt gtcgccatt cgtcagggtc aagctgacga 15360  
actctcagat ataaaagacg tggattacg tgcgttcacg ctgcagggtg gtggtgctga 15420  
attagataaa ctatcgcaat gtgatgtgca aatttatcca gaagcgctat tgaattacaa 15480  
tacgtttgca accgatgaaa aatcattacg ggagatctac cagattggtt atgatgctgc 15540  
aaaagatcaa catgacaacc ttatggcatt gaaagaaagt atcaccacca gcgagggtta 15600  
aaagaacgtc tttagcaaat ggtttgtgta taaactgtct agcaacagcg gcaaatagcg 15660  
gccacacggt atttatacac taggataatg ggcgttaata gcctcactgt cgttgtgtg 15720  
tctctaattt tagctaaatc ttgtgttata ctgacttcct attaatcata aacgatttat 15780  
cacggtaaac atgactcaaa taaataaccc gcttcacggc atgacactcg aaaaagtaat 15840  
taacagtctc gttgaacaat atggctggga tggcttggga tactacatca acattcggtg 15900  
ctttactgaa aatccaagt ttaagtctag tcttaaat ttcgttaaaa ccccttgggc 15960

acgtgataaa gtagaagcgc tatatatcaa aatgggtgact gaaggctaac tgtctccacg 16020  
ctagcgaacc gctgtttata gttaatataa gtactataag cagggtcgt taattcagta 16080  
tgtaattaat cctgaatacc tccgcttatt tcaacattgt actctctaga taacactctc 16140  
aacattacac cttcaacatc acagcctcca cataacatcc gatgacatag ccctgttatt 16200  
tttcacattt atctatatgc tatatatatt agccatttga tcaattgagt taatttctgc 16260  
aatgacaaaag atataccatc atccagtaca aatttattat gaagataccg accattcttg 16320  
tgttgtttac caccctaact ttttaaaata ctttgaacgt gcacgtgagc atgtgataaa 16380  
tagtgactta ctagcaacat tgtggaatga acgcggttta ggttttgagg tgtataaagc 16440  
caatatgact tttcaggatg gggtcgaatt tgctgaagtg tgtgatattc gcacttcttt 16500  
tgtcctagac ggtaagtaca aaacgatctg gcgccaagaa gtatggcgtc cgaatgcgac 16560  
tagggctgcc gttatcgggtg atattgaaat ggtgtgctta gacaaacaaa aacgtttaca 16620  
gcccacccct gatgatgtgt tagctgcaat ggtagtgaa taaatgggtc atgcataaat 16680  
agttaataca tgattctggc ccgtcacgtt tacagataag aggcacccga tgcctccttc 16740  
ctattaccaa tactactgct tatccctttc taactatctt tagcgtccat aacacactga 16800  
gcatttattc tattaatcag tgattgtgat ttaattatct tctatatatg taatttaagt 16860  
taattttcaa tttattttta gctacattaa ggcttacgaa tgtacgctaa aatgagatgt 16920  
cagactaatt ttagcttatt aatctgttag ccgtttatat tttataaaga tgggatttaa 16980  
cttaaatgca attaattatg gcgtaaatag agtgaaaaca tggctaatat tcaactaagtc 17040  
ctgaatttta tataaagttt aatctgttat ttagcgttt acctggctct atcagtgagg 17100  
tttatagcca ttattagtgg gattgaagtg atttttaaag ctatgtatat tattgcaaat 17160  
ataaattgta acaattaaga ctttgacac ttgagttcaa tttcgaattg attggcataa 17220  
aatttaaaac agctaaatct acctcaatca ttttagcaaa tgtatgcagg tagatttttt 17280  
tcgccattta agagtacact tgtacgctag gtttttggtt agtgtgcaaa tgaacgtttt 17340  
gatgagcatt gtttttagag cacaaaatag atccttacag gagcaataac gcaatggcta 17400  
aaaagaacac cacatcgatt aagcacgcca aggatgtgtt aagtagtgat gatcaacagt 17460  
taaattctcg cttgcaagaa tgtccgattg ccatcattgg tatggcatcg gtttttgag 17520  
atgctaaaaa cttggatcaa ttctgggata acatcggtga ctctgtggac gctattattg 17580  
atgtgcctag cgatcgctgg aacattgacg accattactc ggctgataaa aaagcagctg 17640  
acaagacata ctgcaaacgc ggtggtttca ttccagagct tgattttgat ccgatggagt 17700  
ttggtttacc gccaaatata ctcgagttaa ctgacatcgc tcaattgttg tcattaattg 17760  
ttgctcgtga tgtattaagt gatgctggca ttggtagtga ttatgaccat gataaaattg 17820  
gtatcacgct ggggtgcggt ggtggtcaga aacaaatttc gccattaacg tcgcgcctac 17880  
aaggcccggt attagaaaaa gtattaaaag cctcaggcat tgatgaagat gatcgcgcta 17940  
tgatcatcga caaattttaa aaagcctaca tcggctggga agagaactca ttcccaggca 18000  
tgctaggtaa cgttattgct ggctcgatcg ccaatcgttt tgattttggt ggtactaact 18060  
gtgtggttga tgcggcatgc gctggctccc ttgcagctgt taaaatggcg atctcagact 18120  
tacttgaata tcgttcagaa gtcattgatat cgggtggtgt atgttgtgat aactcgccat 18180  
tcatgtatat gtcattctcg aaaacaccag catttaccac caatgatgat atccgtccgt 18240  
ttgatgacga ttcaaaaggc atgctggttg gtgaaggat tggtcatgat gcgtttaaac 18300  
gtcttgaaga tgctgaacgt gacggcgaca aaatttattc tgtactgaaa ggtatcggtg 18360  
catcttcaga tggctggttc aaatctatct acgctccacg cccagatggc caagcaaaaag 18420  
cgctaaaacg tgcttatgaa gatgccggtt ttgccctga aacatgtggt ctaattgaag 18480  
gccatggtac gggtagcaaa gcgggtgatg ccgcagaatt tgctggcttg accaaacact 18540  
ttggcgccgc cagtgatgaa aagcaatata tcgccttagg ctcatgttaa tcgcaaatg 18600  
gtcatactaa atctgcggct ggctctgcgg gtatgattaa ggcggcatta gcgctgcatc 18660  
ataaaatctt acctgcaacg atccatctcg ataaaccaag tgaagccttg gatataaaaa 18720  
acagcccggt atacctaaac agcgaaacgc gtccttggt gccacgtgaa gatggatttc 18780  
cacgtcgtgc aggtatcagc tcatttggtt ttggcgccac caacttccat attattttag 18840



aagagtatcg cccaggtcac gatagcgcat atcgcttaaa ctacgtgagc caaactgtgt 18900  
tgatctcggc aaacgaccaa caaggtattg ttgctgagtt aaataactgg cgtactaaac 18960  
tggctgtcga tgctgatcat caagggtttg tattaatga gttagtgaac acgtggccat 19020  
taaaaacccc atccgttaac caagctcgtt taggtttgtg tgcgcgtaac gcaaatgaag 19080  
cgatcgcgat gattgatacg gcattgaaac aattcaatgc gaacgcagat aaaatgacat 19140  
ggtcagtacc taccggggtt tactatcgtc aagccggtat tgatgcaaca ggtaaagtgg 19200  
ttgcgctatt ctacgggcaa ggctcgcaat acgtgaacat gggctcgtgaa ttaacctgta 19260  
acttcccaaag catgatgcac agtgctgcgg cgatggataa agagttcagt gccgctggtt 19320  
taggccagtt atctgcagtt actttcccta tccctgttta tacggatgcc gagcgtaagc 19380  
tacaagaaga gcaattacgt ttaacgcaac atgcgcaacc agcgattggt agtttgagtg 19440  
ttggtctgtt caaaacgttt aagcaagcag gttttaaacg tgattttgct gccggtcata 19500  
gtttcgggtga gttaaccgca ttatgggctg ccgatgtatt gagcgaaagc gattacatga 19560  
tgtagcgcg tagtcgtggt caagcaatgg ctgcgccaga gcaacaagat tttgatgcag 19620  
gtaagatggc cgctgttggt ggtgatccaa agcaagtcgc tgtgatcatt gatacccttg 19680  
atgatgtctc tattgctaac ttcaactcga ataaccaagt tgttattgct ggtactacgg 19740  
agcaggttgc tgtagcgggt acaaccttag gtaatgctgg tttcaaagtt gtgccactgc 19800  
cggatatctgc tgcgttccat acaccttag ttcgtcacgc gcaaaaacca tttgctaaag 19860  
cggttgatag cgctaaattt aaagcgccaa gcattccagt gtttgctaata ggcacaggct 19920  
tggtgcattc aagcaaaccg aatgacatta agaaaaacct gaaaaaccac atgctggaat 19980  
ctgttcattt caatcaagaa attgacaaca tctatgctga tgggtggcgc gtatttatcg 20040  
aatttggctc aaagaatgta ttaactaaat tgggtgaaa cattctcact gaaaaatctg 20100  
atgtgactgc tatcgcgggt aatgctaata ctaaacacc tgcggacgta caaatgcgcc 20160  
aagctgcgct gcaaatggca gtgcttggtg tcgcattaga caatattgac ccgtacgacg 20220  
ccgttaagcg tccacttggt gcgccgaaag catcaccaat gttgatgaag ttatctgcag 20280  
cgtcttatgt tagtccgaaa acgaagaaag cgtttgctga tgcattgact gatggctgga 20340  
ctgttaagca agcgaaagct gtacctgctg ttgtgtcaca accacaagtg attgaaaaga 20400  
tcgttgaaag tgaaaagata gttgaacgca ttgtcgaaag agagcgtatt gtcgaagtag 20460  
aaaaaatcgt ctacgttaat gctgacgggt cgcttatatc gcaaaaataat caagacgtta 20520  
acagcgctgt tgtagcaac gtgactaata gtcagtgcac tcatagcagt gatgctgacc 20580  
ttgttgctc tattgaacgc agtggtggtc aatttgttgc acaccaacag caattattaa 20640  
atgtacatga acagtttatg caaggtccac aagactacgc gaaaacagtg cagaacgtac 20700  
ttgctgcga gacgagcaat gaattaccg aaagtttaga ccgtacattg tctatgtata 20760  
acgagttcca atcagaaacg ctacgtgtac atgaaacgta cctgaacaat cagacgagca 20820  
acatgaacac catgcttact ggtgctgaag ctgatgtgct agcaacccca ataactcagg 20880  
tagtgaatac agccgttgcc actagtcaca aggtagttgc tccagttatt gctaatacag 20940  
tgacgaatgt tgtatctagt gtcagtaata acgcggcggg tgacgtgcaa actgtggcat 21000  
tagcgcctac gcaagaaatc gctccaacag tcgtactac gccagcacc gcattgggtg 21060  
ctatcgtggc tgaacctgtg attgttgccg atgttgctac agaagttgca ccaattacac 21120  
catcagttac accagttgtc gcaactcaag cggctatcga tgtagcaact attaacaaag 21180  
taatgttaga agttgttgct gataaaaccg gttatccaac ggatattgct gaactgagca 21240  
tggacatgga agctgactta ggtatcgact caatcaaacg tgttgagata ttaggcgcag 21300  
tacaggaatt gatccctgac ttacctgaac ttaactctga agatcttgct gagctacgca 21360  
cgcttggtga gattgtcgat tacatgaatt caaaagccca ggctgtagct cctacaacag 21420  
tacctgtaac aagtgcacct gtttcgcctg catctgctgg tattgattta gccacatcc 21480  
aaaacgtaat gttagaagtg gttgcagaca aaaccgggta cccaacagac atgctagaac 21540  
tgagcatgga tatggaagct gacttaggta ttgattcaat caagcgtgtg gaaatcttag 21600  
gtgcagtaca ggagatcata actgatttac ctgagctaaa ccctgaagat cttgctgaat 21660  
tacgcaccct aggtgaaatc gttagttaca tgcaaagcaa agcgcacgct gctgaaagtg 21720

cgccagtggc gacggctcct gtagcaacaa gctcagcacc gtctatcgat ttgaaccaca 21780  
 ttcaaacagt gatgatggat gtagttgcag ataagactgg ttatccaact gacatgctag 21840  
 aacttggcat ggacatggaa gctgatttag gtatcgattc aatcaaacgt gtggaaatat 21900  
 taggcgcagt gcaggagatc atcactgatt tacctgagct aaacccagaa gacctcgctg 21960  
 aattacgcac gctaggtgaa atcgttagtt acatgcaaag caaagcgcca gtcgctgaga 22020  
 gtgcgccagt agcgacggct tctgtagcaa caagctctgc accgtctatc gatttaaacc 22080  
 atatccaaac agtgatgatg gaagtgggtg cagacaaaac cggttatcca gtagacatgt 22140  
 tagaacttgc tatggacatg gaagctgacc taggtatcga ttcaatcaag cgtgtagaaa 22200  
 ttttaggtgc ggtacaggaa atcattactg acttacctga gcttaaccct gaagatcttg 22260  
 ctgaactacg tacattaggt gaaatcgta gttacatgca aagcaaagcg cccgtagctg 22320  
 aagcgctgc agtacctgtt gcagtagaaa gtgcacctac tagtgtaaca agctcagcac 22380  
 cgtctatcga ttagaccac atccaaaatg taatgatgga tgttggtgct gataagactg 22440  
 gttatcctgc caatatgctt gaattagcaa tggacatgga agccgacctt ggtattgatt 22500  
 caatcaagcg tgttgaaatt ctaggcgcgg tacaggagat cattactgat ttacctgaac 22560  
 taaaccaga agacttagct gaactacgta cgtagaaga aattgtaacc tacatgcaaa 22620  
 gcaaggcgag tgggtgtact gtaaatgtag tggctagccc tgaaaataat gctgtatcag 22680  
 atgcatttat gcaaagcaat gtggcgacta tcacagcggc cgcagaacat aaggcggaat 22740  
 ttaaaccggc gccgagcgca accgttgcta tctctcgtct aagctctatc agtaaaataa 22800  
 gccaagattg taaaggtgct aacgccttaa tcgtagctga tggcactgat aatgctgtgt 22860  
 tacttgaga ccacctattg caaactggct ggaatgtaac tgcattgcaa ccaacttggg 22920  
 tagctgtaac aacgacgaaa gcatttaata agtcagtga cctggtgact taaatggcg 22980  
 ttgatgaaac tgaaatcaac aacattatta ctgtaacgc acaattggat gcagttatct 23040  
 atctgcacgc aagtagcgaa ataatgcta tcgaataccc acaagcatct aagcaaggcc 23100  
 tgaigttagc cttcttatta gcgaaattga gtaaagtaac tcaagccgct aaagtgcgtg 23160  
 gcgcctttat gattgttact cagcaggggtg gttcattagg ttttgatgat atcgattctg 23220  
 ctacaagtca tgatgtgaaa acagacctag tacaaagcgg cttaaacggt ttagttaaga 23280  
 cactgtctca cgagtgggat aacgtattct gtcgtgcggt tgatattgct tcgtcattaa 23340  
 cggctgaaca agttgcaagc cttgttagtg atgaactact tgatgctaac actgtattaa 23400  
 cagaagtggg ttatcaacaa gctggtaaag gccttgaacg tatcacgtta actggtgtgg 23460  
 ctactgacag ctatgcatta acagctggca ataacatcga tgctaactcg gtatttttag 23520  
 tgagtgggtg cgcaaaaggt gtaactgcac attgtgttgc tcgtatagct aaagaatatc 23580  
 agtctaagtt catcttattg ggacgttcaa cgttctcaag tgacgaaccg agctgggcaa 23640  
 gtggtattac tgatgaagcg gcgttaaaga aagcagcgat gcagtcttg attacagcag 23700  
 gtgataaacc aacaccggtt aagatcgtac agctaataca accaatccaa gctaactcgtg 23760  
 aaattgcgca aaccttgtct gcaattaccg ctgctgggtg ccaagctgaa tatgtttctg 23820  
 cagatgtaac taatgcagca agcgtacaaa tggcagtcgc tccagctatc gctaagttcg 23880  
 gtgcaatcac tggcatcatt catggcgcggt gtgtgttagc tgaccaatc attgagcaaa 23940  
 aaactctgag tgatttttag tctgtttaca gcactaaaat tgacggtttg ttatcgctac 24000  
 tatcagtcac tgaagcaagc aacatcaagc aattgggtatt gttctcgtca gcggctggtt 24060  
 tctacggtaa ccccgccag tctgattact cgattgccaa tgagatctta aataaaaccg 24120  
 cataccgctt taaatcattg caccacaag ctcaagtatt gagctttaac tggggctcctt 24180  
 gggacgggtg catggtaacg cctgagctta aacgtatgtt tgaccaacgt ggtgtttaca 24240  
 ttattccact tgatgcaggt gcacagttat tgctgaatga actagccgct aatgataacc 24300  
 gttgtccaca aatcctcgtg ggtaatgact tatctaaaga tgctagctct gatcaaaagt 24360  
 ctgatgaaaa gagtactgct gtaaaaaagc cacaagttag tcgtttatca gatgcttttag 24420  
 taactaaaag tatcaaagcg actaacagta gctctttatc aaacaagact agtgctttat 24480  
 cagacagtag tgcttttcag gttaacgaaa accactttt agctgaccac atgatcaaaag 24540  
 gcaatcaggt attaccaacg gtatgcgcga ttgcttggat gagtgatgca gcaaaagcga 24600

cttatagtaa ccgagactgt gcattgaagt atgtcggttt cgaagactat aaattgttta 24660  
aaggtgtggt ttttgatggc aatgaggcgg cggattacca aatccaattg tcgcctgtga 24720  
caagggcgtc agaacaggat tctgaagtcc gtattgccgc aaagatcttt agcctgaaaa 24780  
gtgacggtaa acctgtgttt cattatgcag cgacaatatt gttagcaact cagccactta 24840  
atgctgtgaa ggtagaactt ccgacattga cagaaagtgt tgatagcaac aataaagtaa 24900  
ctgatgaagc acaagcgta tacagcaatg gcacctgtt ccacggtgaa agtctgcagg 24960  
gcattaagca gatattaagt tgtgacgaca agggcctgct attggcttgt cagataaccg 25020  
atgttgcaac agctaagcag ggatccttcc cgtagctga caacaatatac ttgccaatg 25080  
atttggttta tcaggctatg ttggtctggg tgcgcaaaaca atttggttta ggtagcttac 25140  
cttcgggtgac aacggcttg actgtgtatc gtgaagtggg ttagatgaa gtattttatc 25200  
tgcaacttaa tgttgttgag catgatctat tgggttcacg cggcagtaaa gcccgttgtg 25260  
atattcaatt gattgctgct gatatgcaat tacttgccga agtgaaatca gcgcaagtca 25320  
gtgtcagtga cattttgaac gatatgtcat gatcgagtaa ataataacga taggcgtcat 25380  
ggtagcatg gcgtctgctt tcttcatttt ttaacattaa caatattaat agctaaacgc 25440  
ggttgcttta aaccaagtaa acaagtgtt ttagctatta ctattccaaa caggatatta 25500  
aagagaatat gacggaatta gctgttattg gtatggatgc taaatttagc ggacaagaca 25560  
atattgaccg tgtggaacgc gctttctatg aagggtctta ttagagtaat gttagccgcg 25620  
ttagtaccga atctaattgt attagcaatg gcgaagaaca agttattact gccatgacag 25680  
ttcttaactc tgtcagtcta ctacgcgcaa cgaatcagtt aaatatagct gatatcgcg 25740  
tgttgctgat tgctgatgta aaaagtgtg atgatcagct ttagtccaa attgcatcag 25800  
caattgaaaa acagtgtgcg agttgtgtt tttatgtga tttaggccaa gcattaaatc 25860  
aagtagctga tttagttaat aaccaagact gtcctgtggc tgtaattggc atgaataact 25920  
cggttaattt atctcgtcat gatcttgaat ctgtaactgc aacaatcagc tttgatgaaa 25980  
ccttcaatgg ttataacaat gtagctgggt tcgcgagttt acttatcgct tcaactgcgt 26040  
ttgccaatgc taagcaatgt tatatatacg ccaacattaa gggcttcgct caatcggcg 26100  
taaagtctca atttaacggt gaaacatta gcgatactgc aaagaccgca ttgcagcaag 26160  
ctagcataac tgacagcag gttggtttgt tagaagtgtc agcagtcgct gattcggcaa 26220  
tcgcattgtc tgaaagccaa ggtttaatgt ctgcttatca tcatacgcaa actttgcata 26280  
ctgcattaag cagtgcccg agtggtactg gtgaaggcgg gtgtttttca caggtcgcag 26340  
gtttattgaa atgtgtaatt ggtttacatc aacgttatat tccggcgatt aaagattggc 26400  
aacaaccgag tgacaatcaa atgtcacggt ggcggaattc accattctat atgcctgtag 26460  
atgctcgacc ttggttccca catgctgatg gctctgcaca cattgccgct tatagttgtg 26520  
tgactgctga cagctattgt catattcttt tacaagaaaa cgtcttaca gaacttgtt 26580  
tgaaagaaac agtcttgcaa gataatgact taactgaaag caagcttcag actcttgaa 26640  
aaaacaatcc agtagctgat ctgcgcacta atggttactt tgcacgagc gagttagcat 26700  
taatcatagt acaaggtaat gacgaagcac aattacgctg tgaattagaa actattacag 26760  
ggcagttaag tactactggc ataagtacta tcagtattaa acagatcgca gcagactgtt 26820  
atgcccgtaa tgatactaac aaagcctata gcgcagtgct tattgccgag actgctgaag 26880  
agttaagcaa agaaataacc ttggcgtttg ctggtatcgc tagcgtgttt aatgaagatg 26940  
ctaaagaatg gaaaaccccg aaggcgagtt attttaccgc gcagcctgca aataaacagg 27000  
ctgctaacag cacacagaat ggtgtcacct tcatgtaccc aggtattggg gctacatatg 27060  
ttggtttagg gcgtgatcta tttcatctat tcccacagat ttatcagcct gtagcggctt 27120  
tagccgatga cattggcgaa agtctaaaag atactttact taatccacgc agtattagtc 27180  
gtcatagctt taaagaactc aagcagttgg atctggacct gcgcggtaac ttagccaata 27240  
tcgctgaagc cgggtgtggg tttgcttgtg tgtttacca ggtatttgaa gaagtcttg 27300  
ccgttaaagc tgactttgct acaggttata gcatgggtga agtaagcatg tatgcagcac 27360  
taggctgctg gcagcaaccg ggattgatga gtgctcgct tgcaaatcg aataccttta 27420  
atcatcaact ttgcggcgag ttaagaacac tacgtcagca ttggggcatg gatgatgtag 27480

ctaacggtac gttcgcagcag atctgggaaa cctataccat taaggcaacg attgaacagg 27540  
 tcgaaattgc ctctgcagat gaagatcgtg tgtattgcac cattatcaat acacctgata 27600  
 gcttggtgtt agccggttat ccagaagcct gtcagcgagt cattaagaat ttaggtgtgc 27660  
 gtgcaatggc attgaatatg gcgaacgcaa ttcacagcgc gccagcttat gccgaatacg 27720  
 atcatatggt tgagctatac catatggatg ttactccacg tattaatacc aagatgtatt 27780  
 caagctcatg ttatttaccg attccacaac gcagcaaagc gatttccac agtattgcta 27840  
 aatgtttgtg tgatgtggtg gatttccac gtttggttaa taccttacat gacaaagggtg 27900  
 cgcggttatt cattgaaatg ggtccaggtc gttcggtatg tagctgggtg gataagatct 27960  
 tagttaatgg cgatggcgat aataaaaagc aaagccaaca tgtatctgtt cctgtgaatg 28020  
 ccaaaggcac cagtgatgaa cttacttata ttcgtgcgat tgctaagtta attagtcag 28080  
 gcgtgaattt gaatttagat agcttggtta acgggtcaat cctgggttaa gcaggccata 28140  
 tagcaaacac gaacaaatag tcaacatcga tatctagcgc tggtagtga tacctcatta 28200  
 gttgaaatat ggatttaaag agagtaatta tggaaaatat tgcagtagta ggtattgcta 28260  
 atttgttccc gggctcacia gcaccggatc aattttggca gcaattgctt gaacaacaag 28320  
 attgccgcag taaggcgacc gctgttcaaa tggcggttga tctgtctaaa tataccgcc 28380  
 acaaaggtag cacagataaa ttttactgtg tgcacggcgg ttacatcagt gatttcaatt 28440  
 ttgatgcttc aggttatcaa ctcgataatg attatttagc cggtttagat gaccttaatc 28500  
 aatgggggct ttatgttacg aaacaagccc ttaccgatgc gggttattgg ggcagtactg 28560  
 cactagaaaa ctgtggtgtg attttaggta atttgtcatt cccaactaaa tcatctaatt 28620  
 agctgtttat gcctttgtat catcaagttg ttgataatgc cttaaaggcg gtattacatc 28680  
 ctgattttca attaacgcat tacacagcac cgaaaaaac acatgctgac aatgcattag 28740  
 tagcaggtta tccagctgca ttgatcgcgc aagcggcggg tcttggtggt tcacattttg 28800  
 cactggatgc ggcttggtc tcatctgtt atagcgttaa gttagcgtgt gattacctgc 28860  
 atacgggtaa agccaacatg atgcttgctg gtgcggtatc tgcagcagat cctatgttcg 28920  
 taaatatggg tttctcgata ttccaagctt acccagctaa caatgtacat gccccgttg 28980  
 accaaaattc acaaggctta tttgccgtg aaggcgcggg catgatggta ttgaaacgtc 29040  
 aaagtgatgc agtacgtgat ggtgatcata tttacgccat tattaaaggc ggcgcattat 29100  
 cgaatgacgg taaaggcgag tttgtattaa gccgaacac caagggccaa gtattagtat 29160  
 atgaacgtgc ttatgccgat gcagatgttg acccgagtac agttgactat attgaatgtc 29220  
 atgcaacggg cacacctaa ggtgacaatg ttgaattgcg ttcgatggaa acctttttca 29280  
 gtcgcgtaaa taacaaacca ttactgggct cggttaaatc taaccttggc catttgtaa 29340  
 ctgccgctgg tatgcctggc atgaccaaaag ctatgttagc gctaggtaaa ggtcttattc 29400  
 ctgcaacgat taacttaaag caaccactgc aatctaaaaa cggttacttt actggcgagc 29460  
 aaatgccaac gacgactgtg tcttggtcaa caactccggg tgccaaggca gataaacgc 29520  
 gtaccgcagg tgtgagcgta tttggttttg gtggcagcaa cgccatttg gtattacaac 29580  
 agccaacgca aacactcgag actaatttta gtgttgctaa accacgtgag cctttggcta 29640  
 ttattggtat ggacagccat tttggtatg ccagtaattt agcgcagttc aaaaccttat 29700  
 taaataataa tcaaaatacc ttccgtgaat taccagaaca acgctggaaa ggcattgaaa 29760  
 gtaacgctaa cgtcatgcag tcgttacaat tacgcaaagc gcctaaaggc agttacgttg 29820  
 aacagctaga tattgatttc ttgcgtttta aagtaccgcc taatgaaaaa gattgcttga 29880  
 tcccgaaca gttaatgatg atgcaagtgg cagacaatgc tgcgaaagac ggaggtctag 29940  
 ttgaaggctg taatgttgcg gtattagtag cgatgggcat ggaactggaa ttacatcagt 30000  
 atcgtggctg cgttaatcta accacccaaa ttgaagacag cttattacag caaggtatta 30060  
 acctgactgt tgagcaacgt gaagaactga ccaatattgc taaagacggg gttgcctcgg 30120  
 ctgcacagct aatcagtat acgagtttca ttggtaatat tatggcgtca cgtatttcgg 30180  
 cgttatggga tttttctggt cctgctatta ccgtatcggc tgaagaaaac tctgtttatc 30240  
 gttgtgttga attagctgaa aatctatttc aaaccagtga tgttgaagcc gttattattg 30300  
 ctgctgttga tttgtctggt tcaattgaaa acattacttt acgtcagcac tacgggtccag 30360

ttaatgaaaa gggatctgta agtgaatgtg gtccgggttaa tgaaagcagt tcagtaacca 30420  
 acaatattct tgatcagcaa caatggctgg tgggtgaagg cgacgaggct attgtcggtta 30480  
 aaccgtcatc gcaagtcact gctgagcaag ttatgctggt tattgatgctg gtgagttttg 30540  
 cccctggtag caatgcgaaa gcaattacga ttgcagcggg taaagcatta acacttgctg 30600  
 gtatcagtgct tgctgatgta gctagtgttg aagcacatgc aagtgggttt agtgccgaaa 30660  
 ataatgctga aaaaaccgctg ttaccgactt tatacccaag cgcaagtatc agttcggtga 30720  
 aagccaatat tggtcatacg tttaatgcct cgggtatggc gagtattatt aaaacggcgc 30780  
 tgctgttaga tcagaatacg agtcaagatc agaaaagcaa acatattgct attaacggctc 30840  
 taggtcgtga taacagctgc gcgcatctta tcttatcgag ttcagcgcaa gcgcatcaag 30900  
 ttgcaccagc gcctgtatct ggtatggcca agcaacgccc acagttagtt aaaaccatca 30960  
 aactcgggtg tcagttaatt agcaacgcga ttgttaacag tgcgagttca tctttacacg 31020  
 ctattaaagc gcagtttgcc ggtaagcact taaacaaagt taaccagcca gtgatgatgg 31080  
 ataacctgaa gcccgaaggt attagcgtc atgcaaccaa tgagtatgtg gtgactggag 31140  
 ctgctaacac tcaagcttct aacattcaag catctcatgt tcaagcgtca agtcatgcac 31200  
 aagagatagc accaaaccaa gttcaaaaata tgcaagctac agcagccgct gtaagtccac 31260  
 cccctttctca acatcaacac acagcgcagc ccgtagcggc accgagcgtt gttggagtga 31320  
 ctgtgaaaca taaagcaagt aaccaaattc atcagcaagc gtctacgcat aaagcatttt 31380  
 tagaaagtcg tttagctgca cagaaaaacc tatcgcaact tgttgaattg caaaccaagc 31440  
 tgtaaatcca aactggtagt gacaatacat ctaacaatac tgcgtcaaca agcaatacag 31500  
 tgtaacaaa tctgtatca gcaacgcat taacacttgt gtctaattgc cctgtagtag 31560  
 cgacaaacct aaccagtaca gaagcaaaag cgcaagcagc tgctacacaa gctgggtttc 31620  
 agataaaagg acctgttggt tacaactatc caccgctgca gtttaattgaa cgttataata 31680  
 aaccagaaaa cgtgatttac gatcaagctg atttggttga attcgctgaa ggtgatattg 31740  
 gtaagggtatt tgggtgctgaa tacaatatta ttgatggcta ttcgctcgt gtacgtctgc 31800  
 caacctcaga ttacttgta gtaacacgtg ttactgaact tgatgccaa gtgcatgaat 31860  
 acaagaaatc atacatgtgt actgaatatg atgtgcctgt tgatgcaccg ttcttaattg 31920  
 atggtcagat ccttggtct gttgccgtcg aatcaggcca gtgtgatttg atgttgattt 31980  
 catatatcgg tattgatttc caagcgaaag gcgaacgtgt ttaccgttta cttgattgtg 32040  
 aattaacttt ccttgaagag atggcttttg gtggcgatac ttacgtttac gagatccaca 32100  
 ttgattcgtg tgcacgtaac ggcgagcaat tattattctt cttccattac gattgttacg 32160  
 taggggataa gaaggtactt atcatgcgtg atggttgtgc tggtttcttt actgacgaag 32220  
 aactttctga tggtaaaggc gttattcata acgacaaaga caaagctgag tttagcaatg 32280  
 ctgttaaatc atcattcacg ccgttattac aacataaccg tggtaatac gattataacg 32340  
 acatgatgaa gttggttaat ggtgatgttg ccagttgttt tgggtccgcaa tatgatcaag 32400  
 gtggccgtaa tccatcattg aaattctcgt ctgagaagtt cttgatgatt gaacgtatta 32460  
 ccaagataga cccaaccggt ggtcattggg gactaggcct gttagaaggt cagaaagatt 32520  
 tagaccctga gcattggtat ttcccttgct actttaaaagg tgatcaagta atggctggtt 32580  
 cgttgatgtc ggaaggttgt ggccaaatgg cgatgttctt catgctgtct cttggtatgc 32640  
 ataccaatgt gaacaacgct cgtttccaac cactaccagg tgaatcaca acggtacgtt 32700  
 gtcgtgggca agtactgcca cagcgcaata ccttaactta ccgtatggaa gttactgcga 32760  
 tgggtatgca tccacagcca ttcataaaag ctaatatgta tattttgctt gacggtaaag 32820  
 tgggtgttga tttcaaaaac ttgagcgtga tgatcagcga acaagatgag cattcagatt 32880  
 accctgtaac actgccgagt aatgtggcgc ttaaagcgat tactgcacct gttgcgtcag 32940  
 tagcaccagc atcttcaccc gctaacagcg cggatctaga cgaacgtggt gttgaaccgt 33000  
 ttaagtttcc tgaacgtccg ttaatgcgtg ttgagtcaga cttgtctgca ccgaaaagca 33060  
 aaggtgtgac accgattaag cattttgaag cgcctgctgt tgctggtcat catagagtgc 33120  
 ctaaccaagc accgtttaca ccttggcata tgtttgagtt tgcgacgggt aatatttcta 33180  
 actgtttcgg tcttgatttt gatgtttatg aaggtcgtat tccacctcgt acaccttggt 33240

gcgatttaca agttgttact caggttgtag aagtgcaggg cgaacgtctt gatcttaaaa 33300  
 atccatcaag ctgtgtagct gaatactatg taccggaaga cgcttggtac ttacttaaaa 33360  
 acagccatga aaactggatg ccttatcat taatcatgga aattgcattg caaccaaag 33420  
 gctttatttc tggttacatg ggcacgacgc ttaatacc tgaataagat ctgttcttcc 33480  
 gtaaccttga tggtagcggc acgttattaa agcagattga ttacgcggc aagaccattg 33540  
 tgaataaatc agtcttgggt agtacggcta ttgctgggtg cgcgattatt caaagtttca 33600  
 cgtttgatat gtctgtagat ggcgagctat ttatactgg taaagctgta ttggttact 33660  
 ttagtggtga atcactgact aaccaactgg gcattgataa cggtaaacg actaatgcgt 33720  
 ggtttgttga taacaatacc cccgcagcga atattgatgt gtttgattta actaatcagt 33780  
 cattggctct gtataaagcg cctgtggata aaccgcatta taaattggct ggtggtcaga 33840  
 tgaactttat cgatacagt tcagtgggtg aaggcgggtg taaagcggc gtggcttatg 33900  
 tttatggcga acgtacgatt gatgctgat attggttctt ccgttatcac ttccaccaag 33960  
 atccggtgat gccaggttca ttaggtgtg aagctattat tgagttgat cagacctatg 34020  
 cgcttaaaaa tgatttgggt ggcaagtgt cttaaccacg ttctattgcg ccgatgacgc 34080  
 aagttgattg gaaataccgt gggcaaatta cgcgcgtgaa taaacagat tctactggacg 34140  
 tgcatacac tgagatcgt aatgacgctg gtgaagtgcg aatcgttggt gatgcgaatc 34200  
 tgtctaaaga tggctcgcgt atttatgaag ttaaaaacat cgttttaagt attgttgaag 34260  
 cgtaaaggggt caagtgtaac gtgcttaagc gccgcattgg taaagacgc tttgcacgcc 34320  
 gtgaatccgt ccatggaggc ttgggggttg catccatgcc aacaacagca agcttacttt 34380  
 aatcaatacg gcttggtgtc catttagacg cctcgaactt agtagttaat agacaaaata 34440  
 atttagctgt ggaatgaata tagtaagtaa tcattcggca gctacaaaa aggaattaag 34500  
 aatgtcagat ttaggtttta acaataacaa cgcaattaac tgggcttga aagtagatcc 34560  
 agcgtcagtt catacacaag atgcagaaat taaagcagct ttaatggatc taactaaacc 34620  
 tctctatgtg gcgaataatt caggcgtaac tggatatgt aatcatacgt cagtagcagg 34680  
 tgcgatcagc aataacatcg atgttgatgt attggcgttt gcgcaaaagt taaaccaga 34740  
 agatctgggt gatgatgctt acaagaaaca gcacggcgtt aaatatgctt atcatggcgg 34800  
 tgcgatggca aatggtattg cctcgttga attggtgtt gcgttaggta aagcagggct 34860  
 gttatgttca tttgggtgct caggtctagt gcctgatgcg gttgaagatg caatcgtcg 34920  
 tattcaagct gaattaccaa atggccctta tgcggttaac ttgatccatg caccagcaga 34980  
 agaagcatta gagcgtggcg cggttgaacg tttctaaaa cttggcgtca agacggtaga 35040  
 ggcttcagct taccttgggt taactgaaca cattgtttg tatcgtgctg ctggtctaac 35100  
 taaaaacgca gatggcagtg ttaatatcgg taacaagggt atcgctaaag tatcgcgtac 35160  
 cgaagttggt cgccgcttta tggaaacctgc accgcaaaaa ttactggata agttattaga 35220  
 acaaaataag atcaccctg aacaagctgc tttagcgtt cttgtacctg tggctgatga 35280  
 tattactggg gaagcggatt ctggtggtca tacagataac cgtccgttt taacattatt 35340  
 accgacgatt attggtctgc gtgatgaagt gcaagcgaag tataacttct ctctgcatt 35400  
 acgtgttgggt gctggtggtg gtatcggaac gcctgaagca gactcgtcg catttaacat 35460  
 gggcgcggct tatatcgttc tgggttctgt gaatcaggcg tgtgttgaag cgggtgcatc 35520  
 tgaatatact cgtaaacgtg tatcgacagt tgaaatggct gatgtgacta tggcacctgc 35580  
 tgcagatatg tttgaaatgg gtgtgaagct gcaagtatta aaacgcgggt ctatgttcgc 35640  
 gatgcgtgcg aagaaactgt atgacttgta tgtggcttat gactcgattg aagatatccc 35700  
 agctgctgaa cgtgagaaga ttgaaaaaca aatcttccgt gcaaacctag acgagatttg 35760  
 ggatggcact atcgctttct ttactgaacg cgatccagaa atgctagccc gtgcaacgag 35820  
 tagtctaaa cgtaaatgg cacttatctt ccgttggtat cttggccttt cttcacgctg 35880  
 gtcaaacaca ggcgagaagg gacgtgaaat ggattatcag atttgggcag gcccaagttt 35940  
 aggtgcattc aacagctggg tgaaagggtt ttaccttgaa gactataccc gccgtggcgc 36000  
 tgtagatgtt gctttgcata tgcttaaagg tgctgcgtat ttacaacgtg taaaccagtt 36060  
 gaaattgcaa ggtgttagct taagtacaga attggcaagt tatcgtacga gtgattaatg 36120

ttacttgatg atatgtgaat taattaaagc gcctgagggc gctttttttg gtttttaact 36180  
 cagggtgttg aactcgaaat tgcccccttc aagttagatc gattactcac tcacaatatg 36240  
 ttgatatcgc acttgccata tacttgctca tccaaagccc tatattgata atgggtgttaa 36300  
 tagtctttaa tatccgagtc tttcttcagc ataatactaa tatagagact cgaccaatgt 36360  
 taaacacaac aaagaatata ttcttggtga ctgccttatt attaacgagt gcgagtacga 36420  
 cagctactac gctaaacaat tcgatatcag caattgaaca acgtatttct ggtcgtatcg 36480  
 gtgtggctgt tttagatacg caaaataaac aaacgtgggc ttacaatggg gatgcacatt 36540  
 ttccgatgat gagtacattc aaaaccctcg cttgcgcgaa aatgctaagt gaatcgacaa 36600  
 atggtaatct ggatcccagt actagctcat tgataaaggc tgaagaatta atcccttggg 36660  
 caccagtcac taaaacgttt gtgaataaca ctattacagt ggcgaaagcg tgtgaagcaa 36720  
 caatgctgac cagtataat accgcggcta atattgtttt acagtatatc ggaggccctc 36780  
 aaggcggtac tgcattcttg cgagaaattg gtgatgaaga gagtcagtta gatcgtatag 36840  
 aacctgaatt gaatgaagct aaggctcgag acttgctga taccacgaca ccgaaagcca 36900  
 tagttaccac gctcaacaaa ctactacttg gtgatgttct acttgatttg gataaaaaacc 36960  
 aacttaaaac atggatgcaa aataataaag tgtcagatcc tttactgcgt tctatattac 37020  
 cgcaaggctg gtttattgcc gaccgctcag gtgcgggttg taatggttct cgaggtataa 37080  
 ctgctatgct ttggcactcc gagcgtcaac cgctaatac cagtatttat ttaaccgaaa 37140  
 ctgagttagc aatggcaatg cgcaatgaga ttattgttga gatcggtaag ctgatattca 37200  
 aagaatacgc ggtgaaataa taagttattt ttgataata ctttaacgag cgtagctatc 37260  
 gaagtgaagg cgtcaattag acacctttgc ttcccctaca aaatctaata tgtattacct 37320  
 cggttagtac aattgcccta agttatttct gtccagcttt ggcttagtgc aattgcgtta 37380  
 gccaatgtga acaccaaggg actttgtcgt accataacta ccaagcgact ttgtcgtttt 37440  
 tatcttttct tagacaaaca gaggttaaat gagtgacgcc ttccaaatca caggaatgaa 37500  
 tccgcatttc aataaaatct aaccgctacc aactccgtac aagttgatct ttagttgttt 37560  
 aaaatctata ataaattcaa ttacggaatt aatccgtaca actggaggtt ttatggctac 37620  
 tgcaagactt gatatccgtt tggatgaaga aatcaaagct aaggctgaga aagcatcagc 37680  
 tttactcggc ttaaaaagtt taaccgaata cggtgttcgc ttaatggacg aagattcaac 37740  
 taaagtagtt tctgagcatg agagtattac cggtgaagcg aatgtattcg accaatttat 37800  
 ggctgcttgt gatgaagcga aagcccaaaa taaagcatta cttgaagccg ctgtatttac 37860  
 tcagaatggt gagtttaagt gagttattcc aaacgtttca aagaactgga taaatcaaaa 37920  
 catgacagag catcatttga ctgtggcgaa aaagagctaa atgattttat ccaaactcaa 37980  
 gcagccaaac atatgcaagc aggtattagc cgcactctgy ttttacctgc ttctgcgccg 38040  
 ttaccaaaac aaaaatatcc aatttgctca ttttatagta tcgcgccaaag ctcaattagc 38100  
 cgcgatacgt taccacaagc aatggctaaa aagttaccac gttatcctat ccctgttttt 38160  
 cttttggctc aacttgccgt ccataaagag tttcatggga gtgggttagg caaagttagc 38220  
 ttaattaaag cgttagagta cttttgggaa attaactctc acatgagagc ttacgccatc 38280  
 gttgttgatt gtttaactga acaagctgag tcattctacg ctaaatatgg tttcgacgtt 38340  
 ctctgcgaaa taaatggctg agtaagaatg ttcatatcaa tgaaaacagt caatcagtta 38400  
 ttcacttaac agtaagagtt agtataacag ttgtatgaat taaatttatt atattcggta 38460  
 atctcattgc gatcacgcta gaagtgcgag cgggtcagac cgaggccaca atagcagccg 38520  
 ttacgtttag gggatgactt aaaaagataa ctactacgtc agtggcgatc ctagaggatt 38580  
 aaaggtttat gattcacaac atttatttat tgtgcttaat tttttctatc caatatgcgc 38640  
 aagctgtaaa tatcactgaa gtagactttt atgtcagtga tgatatccct aaagatgttg 38700  
 ccaaattaaa gataggtgaa tccataacga actccagcct tattctaagt aactcatcta 38760  
 ttccactctc gcgggagacg ggtaacatat attactcttc atcaattgct aacttgaact 38820  
 atgactcgat agaatttgtt atggctcaat tgatggccga agattccagc ctttacaaga 38880  
 tgetggtaaa tagcgatagg ttgtccgtgc tagtaatgac atcttccag tccacagatc 38940  
 tctatggctc gacttactcg gcttatttcc ctaatgttgc ggtcatcgat ttgaattgtg 39000

actcgctaac tttagaacat gagctcggcc atctatacgg agctgaacat gaagaaatat 39060  
 atgacgacta tgtcttctat gctgcgatat gtggagacta tacgactatc atgaactcta 39120  
 tgcagcctga aatgaaagaa aaacaaatga taaaggcata ttcattccct gaattaaaag 39180  
 tggatggctt gcagtgcgga aatgaaaata cgaataacaa aaagggttatt ttagacaata 39240  
 ttggtcggtt tagataggat tgggatatta ttctcattcg gctctactta gtgctgttat 39300  
 tatgagtgcc agtgcttcta tctacgatat tggctttaac aagtatttat ctatagacgc 39360  
 taagggtgta tgtatttaag ggatgttcaa gatgaaacta ggtgtaaacg atgtatagtt 39420  
 gtataacatt ttttcaacgg ttggaacgtt cgattctatc gggtaacaag accgcgacga 39480  
 tccgcgataa gtccgatagt cttacttag ttggtcagat gttagatgct tgtactcacg 39540  
 aagataatcg gaaaatgtgt caaatagaaa tactgagcat tgaatatgtg acgttttagtg 39600  
 aattaaaccg tgcgcacgcc aatgctgaag gtttaccgtt tttgtttatg cttaagtggg 39660  
 tagttcgaaa gatattatccg acttcaaagt atttattttt cataagtttc agagtgtgaa 39720  
 ctatcgatat cttataagtc ttagtgcaca aaacagaact atttatagcg ctcaagaagg 39780  
 cgataatttg ataataaatt atgcgcttgt tactattaag agactttaaa tgactgagat 39840  
 ataagatatg acacggaaga acatattgat cacaggcgca agttcagggt tgggccgagg 39900  
 tatggccatc gaatttgcaa aatcagggtc taacttagca ctttgtgcac gtagacttga 39960  
 taatttagtt gcaactgaaa cagaactctt agccctcaat cctcacatcc aaatcgaaat 40020  
 aaaacctctt gatgtcaatg aacatgaaca agtcttcaat gttttccatg aattcaaagc 40080  
 tgaatttggg acgcttgatc gtattattgt taatgctgga ttaggcaagg gtggatcc 40138

&lt;210&gt; 13

&lt;211&gt; 19227

&lt;212&gt; DNA

<213> *Vibrio marinus*

&lt;400&gt; 13

aatgcaatt aattatggcg taaatagagt gaaaacatgg ctaatatcca ctaagtcctg 60  
 aattttatat aaagtttaat ctgttatttt agcgtttacc tggctttatc agtgaggttt 120  
 atagccatta ttagtgggat tgaagtgatt tttaaagcta tgtatattat tgcaaatata 180  
 aattgtaaca attaagactt tggacacttg agttcaattt cgaattgatt ggcataaaat 240  
 ttaaaacagc taaatctacc tcaatcattt tagcaaatgt atgcaggtag atttttttcg 300  
 ccatttaaga gtacacttgt acgctagggt tttgtttagt gtgcaaatga acgttttgat 360  
 gagcattgtt ttagagcac aaaatagatc cttacaggag caataacgca atggctaaaa 420  
 agaacaccac atcgattaag cacgccaagg atgtgttaag tagtgatgat caacagttaa 480  
 attctcgctt gcaagaatgt ccgattgcc aatttggtat ggcacggtt tttgcagatg 540  
 ctaaaaactt ggatcaattc tgggataaca tcgttgactc tgtggacgct attattgatg 600  
 tgcctagcga tcgctggaac attgacgacc attactcggc tgataaaaaa gcagctgaca 660  
 agacatactg caaacgcggg ggtttcattc cagagcttga ttttgatccg atggagtgtg 720  
 gtttaccgcc aaatatcctc gagttaactg acatcgctca attgttgtca ttaattgttg 780  
 ctcgtgatgt attaagtgat gctggcattg gtagtgatta tgaccatgat aaaattggta 840  
 tcacgctggg tgcggtggg ggtcagaaac aaatttcgcc attaacgtcg cgcctacaag 900  
 gcccggtatt agaaaaagta ttaaaagcct caggcattga tgaagatgat cgcgctatga 960  
 tcatcgacaa atttaaaaaa gcctacatcg gctgggaaga gaactcattc ccaggcatgc 1020  
 taggtaacgt tattgctggg cgtatcgcca atcgttttga ttttggtggg actaactgtg 1080  
 tggttgatgc ggcacgctt ggctcccttg cagctgttaa aatggcgatc tcagacttac 1140  
 ttgaatatcg ttcagaagtc atgatcggg gtggtgtatg ttgtgataac tcgccattca 1200  
 tgtatatgtc attctcgaaa acaccagcat ttaccaccaa tgatgatata cgtccgtttg 1260  
 atgacgattc aaaaggcatg ctggttggg aaggtattgg catgatggcg tttaaacgtc 1320



ttgaagatgc tgaacgtgac ggcgacaaaa ttatttctgt actgaaaggt atcggtacat 1380  
cttcagatgg tcgtttcaaa tctatttacg cttcacgccc agatggccaa gcaaaagcgc 1440  
taaaacgtgc ttatgaagat gccggttttg cccctgaaac atgtgggtcta attgaaggcc 1500  
atgggtacggg taccaaagcg ggtgatgccg cagaatttgc tggcttgacc aaacactttg 1560  
gcgcccgcag tgatgaaaag caatatatcg ccttaggctc agttaaatcg caaattgggtc 1620  
atactaaatc tgcggttggc tctgcgggtg tgattaaggc ggcattagcg ctgcatcata 1680  
aaatcttacc tgcaacgatc catatcgata aaccaagtga agccttggat atcaaaaaca 1740  
gcccgttata cctaaacagc gaaacgcgtc cttggatgcc acgtgaagat ggtattccac 1800  
gtcgtgcagg taccagctca tttgggtttg gcggcaccaa cttccatatt attttagaag 1860  
agtatcgccc aggtcacgat agcgcatatc gcttaaaactc agtgagccaa actgtgttga 1920  
tctcggcaaa cgaccaacaa ggtattgttg ctgagttaaa taactggcgt actaaactgg 1980  
ctgtcgatgc tgatcatcaa ggggttctat ttaatgagtt agtgacaacg tggccattaa 2040  
aaaccccatc cgtaaaccaa gctcgtttag gttttgttgc gcgtaatgca aatgaagcga 2100  
tcgcatgat tgatacggca ttgaaacaat tcaatgcgaa cgcagataaa atgacatggt 2160  
cagtacctac cggggtttac tatcgtcaag ccggtattga tgcaacaggt aaagtgggtg 2220  
cgctattctc agggcaaggt tcgcaatacg tgaacatggg tcgtgaatta acctgtaact 2280  
tcccaagcat gatgcacagt gctgcggcga tggataaaga gttcagtgcc gctgggttag 2340  
gccagttatc tgcagttact ttccctatcc ctgtttatac ggatgccgag cgtaagctac 2400  
aagaagagca attacgttta acgcaacatg cgcaaccagc gattggtagt ttgagtgttg 2460  
gtctgttcaa aacgtttaag caagcaggtt ttaaagctga ttttgcgtcc ggtcatagtt 2520  
tcggtgagtt aaccgcatta tgggctgccc atgtattgag cgaaagcgat tacatgatgt 2580  
tagcgcgtag tcgtgggtcaa gcaatggctg cgccagagca acaagatttt gatgcaggta 2640  
agatggcgcg tgtgttgggt gatccaaagc aagtcgctgt gatcattgat acccttgatg 2700  
atgtctctat tgctaacttc aactcgaata accaagttgt tattgctggt actacggagc 2760  
aggttgctgt agcgtttaca accttaggta atgctgggtt caaagttgtg ccactgcccg 2820  
tatctgctgc gtcccataca cctttagttc gtcacgcgca aaaaccattt gctaaagcgg 2880  
ttgatagcgc taaattttaa gcgcgaagca ttccagtgtt tgctaattggc acaggcttgg 2940  
tgcatccaag caaaccgaat gacattaaga aaaacctgaa aaaccacatg ctggaatctg 3000  
ttcatttcaa tcaagaaatt gacaacatct atgctgatgg tggccgcgta tttatcgaat 3060  
ttggtccaaa gaatgtatta actaaattgg ttgaaaacat tctcactgaa aaatctgatg 3120  
tgactgctat cgcggttaat gctaatecta aacaacctgc ggacgtacaa atgcgccaa 3180  
ctgcgctgca aatggcagtg cttgggtgtc cattagacaa tattgacctg tacgacgccg 3240  
ttaagcgtcc acctgttgcg ccgaaagcat caccaatgtt gatgaagtta tctgcagcgt 3300  
cttatgttag tccgaaaacg aagaaagcgt ttgctgatgc attgactgat ggctggactg 3360  
ttaagcaagc gaaagctgta cctgctgttg tgtcacaacc acaagtgatt gaaaagatcg 3420  
ttgaagttga aaagatagtt gaacgcattg tcgaagtaga gcgtattgtc gaagtagaaa 3480  
aaatcgtcta cgtaaatgct gacggttcgc ttatatcgca aaataatcaa gacgttaaca 3540  
gcgctgttgt tagcaacgtg actaatagct cagtgaacta tagcagtgat gctgaccttg 3600  
ttgcctctat tgaacgcagt gttgggtcaat ttgttgaca ccaacagcaa ttattaaatg 3660  
tacatgaaca gtttatgcaa ggtccacaag actacgcgaa aacagtgcag aacgtacttg 3720  
ctgcgcagac gagcaatgaa ttaccgaaa gtttagaccg tacattgtct atgtataacg 3780  
agttccaatc agaaacgcta cgtgtacatg aaacgtacct gaacaatcag acgagcaaca 3840  
tgaacaccat gcttactggg gctgaagctg atgtgctagc aaccccaata actcaggtag 3900  
tgaatacage cgttgccact agtcacaagg tagttgctcc agttattgct aatacagtga 3960  
cgaatgttgt atctagtgtc agtaataacg cggcgggttg agtgcaaaact gtggcattag 4020  
cgcctacgca agaaatcgct ccaacagtcg ctactacgcc agcaccgcga ttggttgcta 4080  
tcgtggctga acctgtgatt gttgcgcatg ttgctacaga agttgcacca attacaccat 4140  
cagttacacc agttgtcgca actcaagcgg ctatcgatgt agcaactatt aacaaagtaa 4200

tgttagaagt tgttgctgat aaaaccggtt atccaacgga tatgctggaa ctgagcatgg 4260  
 acatggaagc tgacttaggt atcgactcaa tcaaacgtgt tgagatatta ggcgcagtac 4320  
 aggaattgat ccctgactta cctgaactta atcctgaaga tcttgctgag ctacgcacgc 4380  
 ttggtgagat tgtcgattac atgaattcaa aagcccaggc tgtagctcct acaacagtac 4440  
 ctgtaacaag tgcacctgtt tcgcctgcat ctgctggtat tgatttagcc cacatccaaa 4500  
 acgtaatggt agaagtgggt gcagacaaaa ccggttacct aacagacatg ctagaactga 4560  
 gcatggatat ggaagctgac ttaggtattg attcaatcaa gcgtgtggaa atcttaggtg 4620  
 cagtacagga gatcataact gatttacctg agctaaaccc tgaagatcct gctgaattac 4680  
 gcaccctagg tgaaatcgtt agttacatgc aaagcaaacg gccagtcgct gaaagtgcgc 4740  
 cagtggcgac ggctcctgta gcaacaagct cagcacctgc tatcgatttg aaccacatte 4800  
 aaacagtgat gatggatgta gttgcagata agactggtta tccaactgac atgctagaac 4860  
 ttggcatgga catggaagct gatttaggtt tcgattcaat caaacgtgtg gaaatattag 4920  
 gcgcagtgca ggagatcatc actgatttac ctgagctaaa ccagaagac ctcgctgaat 4980  
 tacgcacgct aggtgaaatc gttagtaca tgcaaaagca agcgccagtc gctgagagtg 5040  
 cgccagtagc gacggcttct gtagcaacaa gctctgcacc gtctatcgat ttaaaccata 5100  
 tccaaacagt gatgatggaa gtggttgca acaaaaccgg ttatccagta gacatgttag 5160  
 aacttgctat ggacatggaa gctgacctag gtatcgattc aatcaagcgt gtagaaattt 5220  
 taggtgcggt acaggaaatc attactgact tacctgagct taacctgaa gatcttgctg 5280  
 aactacgtac attaggtgaa atcggttagt acatgcaaa caaagcgccc gtagctgaag 5340  
 cgctgcagt acctgttgca gtagaaagt cacctactag tgtaacaagc tcagcacctg 5400  
 ctatcgattt agaccacatc caaatgtaa tgatggatgt tgttgctgat aagactggtt 5460  
 atcctgccaa tatgcttgaa ttagcaatgg acatggaaag cgaccttggg attgattcaa 5520  
 tcaagcgtgt tgaatttcta ggcgcggtac aggagatcat tactgattta cctgaactaa 5580  
 acccagaaga cttagctgaa ctacgtacgt tagaagaaat tgtaacctac atgcaaaagca 5640  
 aggcgagtggt tgttactgta aatgtagtgg ctagccctga aaataatgct gtatcagatg 5700  
 catttatgca aagcaatgtg gcgactatca cagcgggcgc agaacataag gcggaattta 5760  
 aaccggcgcc gagcgcaacc gttgctatct ctcgtctaag ctctatcagt aaaataagcc 5820  
 aagattgtaa aggtgctaac gccctaactg tagctgatgg cactgataat gctgtgttac 5880  
 ttgcagacca cctattgcaa actggctgga atgtaactgc attgcaacca actrgggtag 5940  
 ctgtaacaac gacgaaagca ttaataaagt cagtgaacct ggtgacttta aatggcggtg 6000  
 atgaaactga aatcaacaac attattactg ctaacgcaca attggatgca gttatctatc 6060  
 tgcacgcaag tagcgaaatt aatgctatcg aatacccaca agcatctaag caaggcctga 6120  
 tgttagcctt cttattagcg aaattgagta aagtaactca agccgctaaa gtgcgtggcg 6180  
 cctttatgat tgttactcag cagggtgggt cattaggttt tgatgatatt gattctgcta 6240  
 caagtcatga tgtgaaaaca gacctagtac aaagcggtt aaacggttta gtaagacac 6300  
 tgtctcacga gtgggataac gtattctgtc gtgcggttga tattgcttcg tcattaacgg 6360  
 ctgaacaagt tgcaagcctt gttagtgtg aactacttga tgctaacact gtattaacag 6420  
 aagtgggtta tcaacaagct ggtaaaggcc ttgaacgtat cacgttaact ggtgtggcta 6480  
 ctgacagcta tgcattaaca gctggcaata acatcgatgc taactcggtt tttttagtga 6540  
 gtggtggcgc aaaaggtgta actgcacatt gtgttgctcg tatagctaaa gaatatcagt 6600  
 ctaagttcat cttattggga cgttcaacgt tctcaagtga cgaaccgagc tgggcaagtg 6660  
 gtattactga tgaagcgggc ttaaagaaag cagcgatgca gtctttgatt acagcaggtg 6720  
 ataaaccaac acccggttaag atcgtacagc taatcaaac aatccaagct aatcgtgaaa 6780  
 ttgcgcaaac cttgtctgca attaccgtg ctggtggcca agctgaatat gtttctgcag 6840  
 atgtaactaa tgcagcaagc gtacaaatgg cagtcgctcc agctatcgct aagttcggtg 6900  
 caatcactgg catcattcat ggcgcgggtg tgttagctga ccaattcatt gagcaaaaaa 6960  
 cactgagtga ttttgagttt gtttacagca ctaaaattga cgtttgtta tcgctactat 7020  
 cagtcactga agcaagcaac atcaagcaat cgtattgtt ctcgtcagcg gctgggttct 7080

acggttaaccc cgccagctct gattactcga ttgccaatga gatctttaaataaaaacccgcat 7140  
 accgcttttaa atcattgcac ccacaagctc aagatttgag ctttaactgg ggtccttggg 7200  
 acggtggcat ggtaacgcct gagcttaaac gtatgtttga ccaacgtggt gtttacatta 7260  
 ttccacttga tgcaggtgca cagttattgc tgaatgaact agccgctaata gataaccgtt 7320  
 gtccacaaat cctcgtgggt aatgacttat cttaaagatgc tagctctgat caaaagtctg 7380  
 atgaaaagag tactgctgta aaaaagccac aagttagtcg tttatcagat gcttttagtaa 7440  
 ctaaaagtat caaagcgact aacagtagct ctttatcaaa caagactagt gctttatcag 7500  
 acagtagtgc ttttcagggt aacgaaaacc actttttagc tgaccacatg atcaaaggca 7560  
 atcaggtatt accaacggta tgcgcgattg cttggatgag tgatgcagca aaagcgactt 7620  
 atagtaaccg agactgtgca ttgaagtatg tcggtttcga agactataaa ttgtttaaag 7680  
 gtgtggtttt tgatggcaat gaggcggcgg attaccaaata ccaattgtcg cctgtgacaa 7740  
 gggcgtcaga acaggattct gaagtcgta ttgccgcaaa gatctttagc ctgaaaagtg 7800  
 acggtaaacc tgtgtttcat tatgcagcga caatattgtt agcaactcag ccacttaatg 7860  
 ctgtgaagggt agaacttccg acattgacag aaagtgttga tagcaacaataaagtaactg 7920  
 atgaagcaca agcgttatatac agcaatggca ccttggtcca cggtgaaagt ctgcagggca 7980  
 ttaagcagat attaatgtgt gacgacaagg gcctgctatt ggctgtcag ataaccgatg 8040  
 ttgcaacagc taagcaggga tccttcccgt tagctgacaa caatatcttt gccaatgatt 8100  
 tggtttatca ggctatgttg gtctgggtgc gcaacaataa tggtttaggt agcttacctt 8160  
 cggtgacaac ggcttgact gtgtatcgtg aagtgggtgt agatgaagta ttttatctgc 8220  
 aacttaattgt tgttgagcat gatctattgg gttcacgcgg cagtaaagcc cgttgtgata 8280  
 ttcaattgat tgctgctgat atgcaattac ttgccgaagt gaaatcagcg caagtcagtg 8340  
 tcagtgacat tttgaacgat atgtcatgat cgagtaaata ataacgatag gcgtcatggt 8400  
 gagcatggcg tctgctttct tcatttttta acattaacaa tattaatagc taaacgcggt 8460  
 tgcttttaac caagtaaaca agtgctttta gctattacta ttccaaacag gatattaaag 8520  
 agaatatgac ggaattagct gttattggta tggatgctaa atttagcgga caagacaata 8580  
 ttgaccgtgt ggaacgcgct ttctatgaag gtgcttatgt aggtaattgt agccgcgtta 8640  
 gtaccgaatc taatgttatt agcaatggcg aagaacaagt tattactgcc atgacagttc 8700  
 ttaactctgt cagtctacta gcgcaaacga atcagttaaa tatagctgat atcgcggtgt 8760  
 tgctgattgc tgatgtaaaa agtgctgatg atcagcttgt agtccaaatt gcatcagcaa 8820  
 ttgaaaaaca gtgtgcgagt tgtgtgttta ttgctgattt aggccaagca ttaaatcaag 8880  
 tagctgattt agttaataac caagactgtc ctgtggctgt aattggcatg aataactcgg 8940  
 ttaattttatc tcgtcatgat cttgaatctg taactgcaac aatcagcttt gatgaaacct 9000  
 tcaatggtta taacaatgta gctgggttcg cgagtttact tatcgcttca actgcgtttg 9060  
 ccaatgctaa gcaatgttat atatacgcca acattaagggt cttcgctcaa tcgggcgtaa 9120  
 atgctcaatt taacgttgga aacattagcg atactgcaaa gaccgcattg cagcaagcta 9180  
 gcataactgc agagcagggt ggtttgttag aagtgtcagc agtcgctgat tcggcaatcg 9240  
 cattgtctga aagccaagggt ttaatgtctg cttatcatca tacgcaaaact ttgcatactg 9300  
 cattaagcag tgcccgtagt gtgactgggt aaggcgggtg tttttcacag gtcgcagggt 9360  
 tattgaaatg tgtaattggt ttacatcaac gttatattcc ggcgattaaa gattggcaac 9420  
 aaccgagtga caatcaaagtc tcacgggtgc ggaattcacc attctatatg cctgtagatg 9480  
 ctgcacctg gttcccatat gctgatggct ctgcacacat tgccgcttat agttgtgtga 9540  
 ctgctgacag ctattgtcat attcttttac aagaaaacgt cttacaagaa cttgttttga 9600  
 aagaaacagt cttgcaagat aatgacttaa ctgaaagcaa gcttcagact cttgaacaaa 9660  
 acaatccagt agctgatctg cgcactaatg gttactttgc atcgagcgag ttagcattaa 9720  
 tcatagtaca aggtaatgac gaagcacaat tacgctgtga attagaaact attacaggc 9780  
 agttaagtac tactggcata agtactatca gtattaaaca gatcgagca gactgttatg 9840  
 cccgtaatga tactaacaaa gcctatagcg cagtgttat tgccgagact gctgaagagt 9900  
 taagcaaaga aataacctg gcgtttgctg gtatcgctag cgtgtttaat gaagatgcta 9960

aagaatggaa aaccccgaa ggcagttatt ttaccgcgca gcctgcaa at aaacaggctg 10020  
ctaacagcac acagaatggg gtcaccttca tgtaccagc tattggtgct acatatgttg 10080  
gttttagggcg tgatctatct catctattcc cacagattta tcagcctgta gcggcttttag 10140  
ccgatgacat tggcgaaagt ctaaaagata cttacttaa tccacgcagt attagtcgtc 10200  
atagctttaa agaactcaag cagttggatc tggacctgcg cggtaactta gccaatatcg 10260  
ctgaagcccg tgtgggtttt gcttgtgtgt ttaccaagg atttgaagaa gtctttgccg 10320  
ttaagctga ctttgctaca ggttatagca tgggtgaagt aagcatgtat gcagcactag 10380  
gctgctggca gcaaccggga ttgatgagt ctcgccttgc acaatcgaat accttaatac 10440  
atcaactttg cgcgagttt agaacactac gtcagcattg gggcatggat gatgtagcta 10500  
acggtagctt cgagcagatc tgggaaacct ataccattaa ggcaacgatt gaacaggctg 10560  
aaattgcctc tgcagatgaa gatcgtgtgt attgcacat tatcaatata cctgatagct 10620  
tgttgtagc cggttatcca gaagcctgtc agcgagtc atagaattta ggtgtgcgtg 10680  
caatggcatt gaatatggcg aacgcaattc acagcgcgcc agcttatgcc gaatacgc 10740  
atatggttga gctataccat atggatgtta ctccacgtat taataccaag atgtattcaa 10800  
gctcatgtta tttaccgatt ccacaacgca gcaagcgat tccccacagt attgctaaat 10860  
gtttgtgtga tgtggtggat tccccaggt tggtaatac cttacatgac aaagggtgcg 10920  
gggtattcat tgaatgggt ccaggtcgtt cggtatgtag ctgggtagat aagatcctag 10980  
ttaatggcg tggcgataat aaaaagcaaa gccaacatgt atctgttct gtgaatgcc 11040  
aaggcaccag tgatgaactt acttatattc gtgcgattgc taagttaatt agtcatggcg 11100  
tgaatttgaa tttagatagc ttgtttaacg ggtcaatcct ggttaaagca ggcatatag 11160  
caaacacgaa caaatagtca acatcgatat ctacgcgtg tgagttatac ctcattagtt 11220  
gadatatgga tttaaagaga gtaattatg aaaatattgc agtagtaggt attgctaatt 11280  
tgttcccggt ctcacaagca cggatcaat tttggcagca attgcttgaa caacaagatt 11340  
gccgcagtaa ggcgaccgct gttcaaatg gcgttgatcc tgctaaatat accgccaaca 11400  
aaggtagcac agataaattt tactgtgtgc acggcggtta catcagtgtt ttcaattttg 11460  
atgcttcagg ttatcaactc gataatgatt atttagccgg tttagatgac cttaatcaat 11520  
gggggcttta tgttacgaaa caagccctta cggatgcggg ttattggggc agtactgcac 11580  
tagaaaactg tgggtgtgatt ttaggtaatt tgtcattccc aactaaatca tctaatacgc 11640  
tgtttatgcc tttgtatcat caagttgtt ataatgcctt aaaggcggtta ttacatcctg 11700  
attttcaatt aacgcattac acagcaccga aaaaaacaca tgcagacaat gcattagtag 11760  
caggttatcc agctgcattg atcgcgcaag cggcggttct tgggtggttca cttttgcac 11820  
tggatgcggc ttgtgcttca tctgttata gcgttaagt agcgtgtgat tacctgcata 11880  
cgggtaaagc caacatgat cttgctggtg cggatctgc agcagatcct atgttcgtaa 11940  
atatgggttt ctcgatattc caagcttacc cagctaaca tgtacatgcc ccgtttgacc 12000  
aaaattcaca aggtctatct gccggtgaag gcgcgggcat gatggtattg aaacgtcaaa 12060  
gtgatgcagt acgtgatggt gatcatattt acgccattat taaaggcggc gcattatcga 12120  
atgacggtaa aggcgagttt gtattaagcc cgaacaccaa gggccaagta ttagtatatg 12180  
aacgtgctta tgccgatgca gatgttgacc cagtagcagt tgactatatt gaatgtcatg 12240  
caacgggcac acctagggt gacaatgtt aattgcgtt gatggaaacc tttttcagtc 12300  
gcgtaaataa caaaccatta ctgggctcgg ttaaatctaa ccttggtcat ttgttaactg 12360  
ccgtggtat gcctggcatg accaaagcta tggtagcgt aggtaaagg cttattcctg 12420  
caacgattaa cttaaagcaa ccaactgcaat ctaaaaacgg ttactttact ggcgagcaaa 12480  
tgccaacgac gactgtgtct tggccaaca ctccgggtgc caaggcagat aaaccgcgta 12540  
ccgcaggtgt gagcgtatct ggttttgggt gcagcaacgc ccatttggtt ttacaacagc 12600  
caacgcaaac actcgagact aattttagt ttgctaaacc acgtgagcct ttggctatta 12660  
ttggtatgga cagccatttt gtagtgcca gtaatttagc gcagttcaaa acctatttaa 12720  
ataataatca aaataccttc cgtgaattac cagaacaacg ctggaaaggc atggaaagta 12780  
acgctaacgt catgcagtcg ttacaattac gcaagcgcc taaaggcagt tacgttgaac 12840

agctagatat tgatttcttg cgttttaaaag taccgcctaa tgaaaaagat tgcttgatcc 12900  
 cgcaacagtt aatgatgatg caagtggcag acaatgctgc gaaagacgga ggtctagttg 12960  
 aaggtcgtaa tgttgcggtg ttagtagcga tgggcatgga actggaatta catcagatc 13020  
 gtggtcgcgt taatctaacc acccaaattg aagacagctt attacagcaa ggtattaacc 13080  
 tgactgttga gcaacgtgaa gaactgacca atattgctaa agacgggtgt gcctcggctg 13140  
 cacagctaaa tcagtatacg agtttcattg gtaatattat ggcgtcacgt atttcggcgt 13200  
 tatgggattt ttctggctct gctattaccg tatcggctga agaaaactct gtttatcgtt 13260  
 gtgttgaaat agctgaaaat ctatttcaaa ccagtgatgt tgaagccgtt attattgctg 13320  
 ctgttgattt gtctggttca attgaaaaca ttactttacg tcagcactac ggtccagtta 13380  
 atgaaaaggg atctgtaagt gaatgtggtc cgggttaatga aagcagttca gtaaccaaca 13440  
 atattcttga tcagcaacaa tggctggtgg gtgaaggcgc agcggctatt gtcgttaaac 13500  
 cgtcatcgca agtcactgct gagcaagttt atgcgcgtat tgatgcgggt agttttgccc 13560  
 ctggtagcaa tgcgaaagca attacgattg cagcggataa agcattaaca cttgctggta 13620  
 tcagtgtgc tgatgtagct agtgttgaag cacatgcaag tggttctagt gccgaaaata 13680  
 atgctgaaaa aaccgcgtta ccgactttat acccaagcgc aagtatcagt tcggtgaaag 13740  
 ccaatattgg tcatacgttt aatgcctcgg gtatggcgag tattattaaa acggctcag 13800  
 tgttagatca gaatacaggt caagatcaga aaagcaaaca tattgctatt aacggtcag 13860  
 gtcgtgataa cagctgcgcg catcttatct tatcgagttc agcgaagcg catcaagttg 13920  
 caccagcgcc tgtatctggt atggccaagc aacgcccaca gttagttaaa accatcaaac 13980  
 tcggtggtca gtttaattagc aacgcgattg ttaacagtgc gagtccatct ttacacgcta 14040  
 ttaaagcgca gtttgccggt aagcacttaa acaaagttta ccagccagtg atgatggata 14100  
 acctgaagcc ccaagggtatt agcgtcatg caaccaatga gtatgtggtg actggagctg 14160  
 ctaacactca agcttctaac attcaagcat ctcatgttca agcgtcaagt catgcacaag 14220  
 agatagcacc aaaccaagtc caaaatatgc aagctacagc agccgctgta agttcacccc 14280  
 ttctccaaca tcaacaacga gcgcagcccg tagcggcacc gagcgttgtt ggagtgaactg 14340  
 tgaaacataa agcaagtaac caaattcatc agcaagcgtc tacgcataaa gcatttttag 14400  
 aaagtctgtt agctgcacag aaaaacctat cgcaacttgt tgaattgcaa accaagctgt 14460  
 caatccaaac tggtagtgac aatacatcta acaatactgc gtcaacaagc aatacagtc 14520  
 taacaaatcc tgtatcagca acgccattaa cacttgtgtc taatgcgcct gtagtagcga 14580  
 caaacctaac cagtacagaa gcaaaagcgc aagcagctgc tacacaagct ggttttcaga 14640  
 taaaaggacc tgttgggtac aactatccac cgctgcagtt aattgaacgt tataataaac 14700  
 cagaaaacgt gatttacgat caagctgatt tggttgaatt cgctgaagggt gatattggtg 14760  
 aggtatttgg tgttgaatac aatattattg atggctattc gcgtcgtgta cgtctgccaa 14820  
 cctcagatta cttgttagta acacgtgtta ctgaacttga tgccaagggt catgaataca 14880  
 agaaatcata catgtgtact gaatatgat tgcctgttga tgcaccgttc ttaattgatg 14940  
 gtcagatccc ttggtctgtt gccgtcgaat caggccagtg tgatttgatg ttgatttcat 15000  
 atatcggat tgatttccaa gcgaaaggcg aacgtgttta ccgtttactt gattgtgaat 15060  
 taactttcct tgaagagatg gcttttggtg gcgatacttt acgttacgag atccacattg 15120  
 attcgtatgc acgtaacggc gagcaattat tattcttctt ccattacgat tgttacgtag 15180  
 gggataagaa ggtacttata atgcgtaatg gttgtgctgg ttcttttact gacgaagaac 15240  
 ttcttgatgg taaaggcgtt attcataacg acaaagacaa agctgagttt agcaatgctg 15300  
 ttaaatcatt attcacgccg ttattacaac ataaccgtgg tcaatacgat tataacgaca 15360  
 tgatgaagtt ggtraatggt gatgttgcca gttgttttgg tccgcaatat gatcaagggtg 15420  
 gccgtaatcc atcattgaaa ttctcgtctg agaagttctt gatgattgaa cgtattacca 15480  
 agatagaccc aaccggtggt cattggggac taggcctgtt agaaggtcag aaagatttag 15540  
 accctgagca ttggtatttc cttgttact ttaaagggtg tcaagtaatg gctggttcgt 15600  
 tgatgtcggg aggttgtggc caaatggcga tgttcttcat gctgtctctt ggtatgcata 15660  
 ccaatgtgaa caacgctcgt ttccaaccac taccaggtga atcacaacag gtacgtctgc 15720

gtgggcaagt actgccacag cgcaatacct taacttaccg tatggaagtt actgcatgg 15780  
 gtatgcatcc acagccattc atgaaagcta atattgatat ttgtcttgac ggtaaaagtg 15840  
 ttgttgattt caaaaaacttg agcgtgatga tcagcgaaca agatgagcat tcagattacc 15900  
 ctgtaacact gccgagtaat gtggcgctta aagcgattac tgcacctgtt gcgtcagtag 15960  
 caccagcatc ttcacccgct aacagcgcggt atctagacga acgtggtgtt gaaccgttta 16020  
 agtttcttga acgtccgtta atgcgtgttg agtcagactt gtctgcaccg aaaagcaaa 16080  
 gtgtgacacc gattaagcat ttgaaagcgc ctgctgttgc tggatcatcat agagtgccta 16140  
 accaagcacc gtttacacct tggcatatgt ttgagtttgc gacgggtaat atttctaact 16200  
 gtttcggtcc tgattttgat gtttatgaag gtcgtattcc acctcgtaca ctttgtggcg 16260  
 atttacaagt tgttactcag gttgtagaag tgcagggcga acgtcttgat cttaaaaatc 16320  
 catcaagctg tgtagctgaa tactatgtac cggaagacgc ttggtacttt actaaaaaca 16380  
 gccatgaaaa ctggatgctt tattcattaa tcattggaat tgcattgcaa ccaaatggct 16440  
 ttatttcttg ttacatgggc acgacgctta aataccctga aaaagatctg ttcttccgta 16500  
 accttgatgg tagcggcacg ttattaaagc agattgattt acgcggaag accatttgtga 16560  
 ataaatcagt cttggctagt acggctattg ctgggtggcg gattattcaa agtttcacgt 16620  
 ttgatatgtc tgtagatggc gagctatttt atactggtaa agctgtattt gggtacttta 16680  
 gtggtgaatc actgactaac caactgggca ttgataacgg taanaacgact aatgcgtgg 16740  
 ttgttgataa caataccccc gcagcgaata ttgatgtgtt tgatttaact aatcagtcac 16800  
 tggctcttga taaagcgctt gtggataaac cgcattataa attggctggg ggtcagatga 16860  
 actttatcga tacagtgtca gtgggtgaag gcggtggtaa agcgggctgt gcttatgttt 16920  
 atggcgaacg tacgattgat gctgatgatt gttcttccg ttatcacttc caccaagatc 16980  
 cgggtgatgc aggttcatta ggtgttgaag ctattattga gttgatgcag acctatgcgc 17040  
 ttaaaaaatga tttgggtggc aagtttgcga acccacgttt cattgcgccg atgacgcaag 17100  
 ttgattgga ataccgtggg caaattacgc cgtggaataa acagtgtca ctggacgtgc 17160  
 atactactga gatcgtgaat gacgctgggt aagtgcgat cgttgggtgat gcgaatctgt 17220  
 cttaaagatg tctgcgtatt tatgaagtta aaaacatcgt tttaagtatt gtrgaagcgt 17280  
 aaaggggtcaa gtgtaacgtg cttaaagcgc gcattggtta aagacgcttt gcacgccgtg 17340  
 aatccgtcca tggaggcttg gggttggcat ccattgccac aacagcaagc ttactttaat 17400  
 caatacggct tgggtgtccat tttagacgct cgaacttagt agttaataga caaaaataat 17460  
 tagctgtgga atgaatatag taagttaatca ttcggcagct acaaaaaagg aattaagaat 17520  
 gtcgagttta ggttttaaca ataacaacgc aattaactgg gcttggaaaag tagatccagc 17580  
 gtcagttcat acacaagatg cagaaattaa agcagcttta atggatctaa ctaaaacctt 17640  
 ctatgtggcg aataattcag gcgttaactgg tatagctaata catacgtcag tagcaggtgc 17700  
 gatcagcaat aacatcgatg ttgatgtatt ggcgtttgcg caaaagttaa acccagaaga 17760  
 tctgggtgat gatgcttaca agaaacagca cggcggttaa tatgcttatc atggcggtgc 17820  
 gatggcaaat ggtattgcct cgggtgaatt ggtgttgcg ttaggtaaaag cagggtgtt 17880  
 atgttcattt ggtgctgcag gtctagtgc tgatgcggtt gaagatgcaa ttcgtcgtat 17940  
 tcaagctgaa ttaccaaagt gcccttatgc ggttaacttg atccatgcac cagcagaaga 18000  
 agcattagag cgtggcgcggt ttgaacgttt cctaaaactt ggcgtcaaga cggtagaggc 18060  
 ttcagcttac cttggtttta ctgaacacat tgtttggtat cgtgctgctg gtctaactaa 18120  
 aaacgcagat ggcagtgtta atatcggtta caaggttctc gctaaagtat cgcgtaccga 18180  
 agttggctgc cgttttatgg aacctgcacc gcaaaaatta ctggataagt tattagaaca 18240  
 aaataagatc acccctgaac aagctgcttt agcgttgctt gtacctatgg ctgatgatat 18300  
 tactggggaa gcggattctg gtggtcatac agataaccgt ccgtttttta cattattacc 18360  
 gacgattatt ggtctgcgtg atgaagtgc agcgaagat aacttctctc ctgcattacg 18420  
 tgttgggtgt ggtgggtgta tcggaacgcc tgaagcagca ctgctgcat ttaacatggg 18480  
 cgcggcttat atcgttcttg gtctcttgaa tcaggcgtgt gtrgaagcgg gtgcatctga 18540  
 atataactct aaactgttat cgacagttga aatggctgat gtgactatgg cacctgctgc 18600

```

agatatgttt gaaatgggtg tgaagctgca agtattaaaa cgcggttcta tgttcgcgat 18660
gcgtgcgaag aaactgtatg acttgatatg ggcttatgac tcgattgaag atatcccagc 18720
tgctgaacgt gagaagattg aaaaacaaat cttccgtgca aacctagacg agatttggga 18780
tggcactatc gctttcttta ctgaacgcga tccagaaatg ctagcccgtg caacgagtag 18840
tcctaaacgt aaaatggcac ttatcttccg ttggtatctt ggcctttctt cacgctggtc 18900
aaacacaggc gagaagggac gtgaaatgga ttatcagatt tgggcaggcc caagtttagg 18960
tgcattcaac agctgggtga aaggttctta ccttgaagac tataaccgcc gtggcgctgt 19020
agatgttgct ttgcatatgc ttaaaggcgc tgcgtattta caacgtgtaa accagttgaa 19080
attgcaaggt gttagcttaa gtacagaatt ggcaagttat cgtacgagtg attaagttaa 19140
cttgatgata tgtgaattaa ttaaagcgcc tgaggcgct ttttttggtt ttttaactcag 19200
gtgttgtaac tcgaaattgc ccctttc 19227

```

&lt;210&gt; 14

&lt;211&gt; 217

&lt;212&gt; DNA

<213> *Shewanella putrefaciens*

&lt;400&gt; 14

```

attggtaaaa ataggggtta tgtttggtgc tttaaagagt gtcctgaaaa attgctaact 60
tctcgattga tttccttata cttctgtccg ttaacaatac aagagtgcga taaccagact 120
acagagttgg ttaagtcatt gctgcctgaa gatgagttaa ttaagggttaa tcgctacatt 180
aaacaagaag ctaaaactca aggtttaatg gtaagag 217

```

&lt;210&gt; 15

&lt;211&gt; 72

&lt;212&gt; PRT

<213> *Shewanella putrefaciens*

&lt;400&gt; 15

Ile Gly Lys Asn Arg Gly Tyr Val Cys Cys Phe Lys Glu Cys Pro Glu

1

5

10

15

Lys Leu Leu Thr Ser Arg Leu Ile Ser Leu Tyr Phe Cys Pro Leu Thr

20

25

30

Ile Gln Glu Cys Asp Asn Gln Thr Thr Glu Leu Val Lys Ser Trp Leu

35

40

45

Pro Glu Asp Glu Leu Ile Lys Val Asn Arg Tyr Ile Lys Gln Glu Ala

50

55

60

Lys Thr Gln Gly Leu Met Val Arg

65

70

&lt;210&gt; 16

&lt;211&gt; 885

&lt;212&gt; DNA

<213> *Shewanella putrefaciens*

<400> 16

```

agcgaaatgc ttatcaagaa attccaagat caatacatca ctgggaagaa aattcattcc 60
ctgggttcaact gggtaacgtt atttccggcc gtattgctaa ccgcttcgac cttggtggca 120
tgaactgtgt cgttgatgca gcatgtgcag gccctcttgc tgcattgcgt atggcattaa 180
gcgagcttgt tgaaggccgc agcgaaatga tgattacagg tgggtgtgt accgataact 240
caccaaccat gtacatgagc ttctctaaaa caccggcatt caccgacaaac gaaacaattc 300
aaccattcga tattgactcg aaaggtatga tgattggtga aggtatcggg atgattgcgc 360
ttaaacgtct tgaagacgca gacgtgatg gcgaccgtat ctattccgtg attaaaggtg 420
ttgggtgcat cttcagacgg taatttatta agagtantta tgcgntcgt cctgaaggtc 480
aggctaaggc acttaaactg gttacgacg atgcagggtt cgcaccgcac acacttggct 540
tacttgaagc ccacggcaca ggcacagcag caggtgatgt ggcagaattc agtggcttta 600
actctgtatt cagtgaaggc aatgacgaaa agcaacacat cgcattaggt tcagtgaat 660
cacagattgg tcacactaaa tcaacagcgg gtactgcggg tctaataaaa gcgtctttag 720
cactgcacca taaagtactg ccgccaacaa tcaatgtaac cagccctaac cctaaactga 780
atattgaaga ctgcctttc tacctcaata cacagacgcg tccatggatg caacgtgtcg 840
atggtacacc gcgtcgtgct ggtattagct catttgggtt tgggtg 885

```

<210> 17

<211> 409

<212> DNA

<213> *Shewanella putrefaciens*

<400> 17

```

ccaagctaaa gcacttaacc gtgcttatga agatgccggt tttgccctg aaacatgtgg 60
tctaattgaa ggccatggta cgggtaccaa agcgggtgat gccgcagaat ttgctggtt 120
gaccaaacac tttggcgccg ccagtgatga aaagcaatat atcgcccttag gctcagttaa 180
atcgcaaat ggtcatacta aatctgcggc tggctctgcg ggtatgatta agcggcatt 240
agcgtgcat cataaaatct tacctgcaac gatccatctc gataaaccaa gtgaagcctt 300
ggatatcaaa aacagcccgt tatacctaaa cagcgaaacg cgtccttgga tgccacgtga 360
agatggtatt ccacgtcgtg caggtattag ctcatctggt tttggtggc 409

```

<210> 18

<211> 81

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 18

```

ccaagctaaa gcacttaacc gtgcctatga tgatgccggt tttgccctg aaacatgtgg 60
tctaattgaa ggccatggta c
81

```

<210> 19

<211> 81

<212> DNA



<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 19

ccaagctaaa gcacttaacc gtgcttatga agatgccggt ttgcccctg aaacatgtgg 60  
tctaattgaa ggccatggta c 81

<210> 20

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 20

agaacgcaaa gttgccgcac tgtttggtcg ccaaggttca caa 43

<210> 21

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 21

caaagcgggt gatgccgcac tgtttggtcg cttgacctaa cac 43

<210> 22

<211> 55

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 22

cattgcgcta ggttcagtta aatcacaaat tggtcatact aaatcaactg caggt 55

<210> 23

<211> 55

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 23

tatcgccctta ggctcagtta aatcgcaaat tggtcatact aaatctgcgg ctggc 55

<210> 24

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 24

cggttcgat tttggcggca tgaacggtg 29

<210> 25

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 25

cgcgatgat taaggcggca ttagcgctg 29

<210> 26

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 26

gcactgctgc aagcatgaac gcgtcggt 28

<210> 27

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 27

gctctgcggc tatcattaac gcggcatt 28

<210> 28

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 28

tccctggtgc taaccatatc agcaaacca

29

<210> 29

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 29

tacctgcaac gatccatatc gataaacca

29

<210> 30

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 30

ctcacctttg tatctaaaca ctgagacttc gtccatgggt accacgtggt gatggtacgc 60  
cgcgccgcgc gggtagtagc tcatttggtt ttggtggc 98

<210> 31

<211> 98

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: SYNTHETIC

<400> 31

cagcccggtta tacctaaaca gcgaaacggc gtccttggat gccacgtgaa gatggtattc 60  
cacgtcgtgc aggtatttagc tcatttggtt ttggtggc 98

<210> 32

<211> 4  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 32  
Asp Xaa Ala Cys  
1

<210> 33  
<211> 4  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 33  
Gly Phe Gly Gly  
1

<210> 34  
<211> 5  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 34  
Gly His Ser Xaa Gly  
1 5

<210> 35  
<211> 6  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 35  
Leu Gly Xaa Asp Ser Leu  
1 5

<210> 36  
<211> 6  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 36  
Leu Gly Xaa Asp Ser Ile  
1 5

<210> 37  
<211> 6  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 37  
Gly Xaa Gly Xaa Xaa Gly  
1 5

<210> 38  
<211> 6  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 38  
Gly Xaa Gly Xaa Xaa Ala  
1 5

<210> 39  
<211> 6  
<212> PRT  
<213> 'Axial Seamount' polynoid polychaete

<400> 39  
Gly Xaa Gly Xaa Xaa Pro  
1 5

<210> 40  
<211> 5  
<212> PRT  
<213> *Shewanella putrefaciens*

<400> 40  
Gly Xaa Ser Xaa Gly  
1 5

<210> 41  
<211> 35  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 41  
cuacuacuac uaccaagcta aagcacttaa ccgtg 35

<210> 42  
<211> 32  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 42  
cuacuacuac uaacagcgaa atgcttatca ag 32

<210> 43  
<211> 38  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 43  
cuacuacuac uagcgaccaa aaccaaata gctaatac 38

<210> 44  
<211> 12  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 44  
aagcccgggc tt 12

<210> 45  
<211> 20  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 45  
gtacaagccc gggcttagct 20

<210> 46

<211> 56

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 46

cgcgatttaa atggcgcgcc ctgcaggcgg ccgcctgcag ggcgcgccat taaat 56

<210> 47

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 47

ctgcagctcg agacaatgtt gatttcctta tacttctgtc c 41

<210> 48

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 48

ggatccagat ctctagctag tcttagctga agctcga 37

<210> 49

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 49

tctagactcg agacaatgag ccagacctct aaacctaca 39

<210> 50

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 50

cccgggctcg agctaattcg cctcactgtc gtttgct

37

<210> 51

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 51

gaattctctg agacaatgcc gctgcgcata gcacttatc

39

<210> 52

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 52

ggtaccagat cttagactt ccccttgaag taaatgg

37

<210> 53

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 53

gaattcgctg acacaatgac attaccagac aatgcttct

39

<210> 54

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic



<400> 54  
tctagagtcg acttatacag attcttcgat gctgatatg 38

<210> 55  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 55  
gaattcgtcg acacaatgaa tcctacagca actaacgaa 39

<210> 56  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 56  
tctagaggat ccttaggccca ttctttggtt tggcttc 37

<210> 57  
<211> 39  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 57  
tctagagtcg acacaatggc ggaattagct gttattggt 39

<210> 58  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: synthetic

<400> 58  
gtcgacggat ccctatttgt tcgtgtttgc tatatg 36

<210> 59

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 59

gtcgacggat ccacaatgaa tatagtaagt aatcattcgg ca

42

<210> 60

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 60

gtcgacctcg agttaatcac tcgtacgata acttgcc

37

<210> 61

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 61

cccgggtcga cacaatggct aaaaagaaca ccacatcga

39

<210> 62

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: synthetic

<400> 62

cccgggtcga ctcatgacat atcgttcaaa atgtcactga

40

<210> 63

<211> 44

<212> DNA

<213> Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: synthetic

&lt;400&gt; 63

tcgacatgga aaatattgca gtagtaggta ttgctaattt gttc

44

&lt;210&gt; 64

&lt;211&gt; 44

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: synthetic

&lt;400&gt; 64

ccgggaacaa attagcaata cctactactg caatattttc catg

44

&lt;210&gt; 65

&lt;211&gt; 21

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: synthetic

&lt;400&gt; 65

tcagatgaac tttatcgata c

21

&lt;210&gt; 66

&lt;211&gt; 36

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: synthetic

&lt;400&gt; 66

tcatgagacg tcgtcgactt acgcttcaac aatact

36

&lt;210&gt; 67

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 67

gtgatgatct ttccctgatg cacgccaagg

30

&lt;210&gt; 68

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 68

agctcgagac cggcaacccg cagcgccaga

30

&lt;210&gt; 69

&lt;211&gt; 4446

&lt;212&gt; DNA

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 69

cgctgccgcc gcgtctcgcc gcgccgcgcc gcgccgcgcc cgccgctcgc gcgcacgccc 60  
gcgcgtctcg ccgcgcctgc tgtctcgaac gagcttctcg agaaggccga gaccgtcgtc 120  
atggaggtcc tcgccgccaa gactggctac gagactgaca tgatcgagtc cgacatggag 180  
ctcgagactg agctcggcat tgactccatc aagcgtgtcg agatcctctc cgaggttcag 240  
gccatgctca acgtcgaggc caaggacgtc gacgctctca gccgcactcg cactgtgggt 300  
gaggtcgtca acgccatgaa ggctgagatc gctgggtggt ctgccccggc gcctgccgcc 360  
gctgccccag gtccggctgc tgccgccctc gcgcctgtcg tctcgagcga gcttctcgag 420  
aaggccgaga ctgtcgtcat ggaggtctct gccgccaaaga ctggctacga gactgacatg 480  
attgagtccg acatggagct cgagaccgag ctcggcattg actccatcaa gcgtgtcgag 540  
attctctccg aggttcaggc catgctcaac gtcgaggcca aggacgtcga cgctctcagc 600  
cgcactcgca ctgttggtga ggtcgtcgat gccatgaagg ctgagatcgc tggcagctcc 660  
gcctcggcgc ctgccgcgcc tgctcctgct ccggtgtgtg ccgctcctgc gcccgctgcc 720  
gccgccccctg ctgtctcgaa cgagcttctc gagaaagccg agactgtcgt catggaggtc 780  
ctcgccgccca agactggcta cgagactgac atgatcgagt ccgacatgga gctcgagact 840  
gagctcggca ttgactccat caagcgtgtc gagatcctct ccgagggtca ggccatgtc 900  
aacgtcgagg ccaaggacgt cgatgccctc agccgcaccc gcactgttgg cgaggttgtc 960  
gatgccatga aggccgagat cgctgggtggc tctgccccgg cgcctgccgc cgctgcccct 1020  
gctccggctg ccgccgcccc tgctgtctcg aacgagcttc ttgagaaggc cgagactgtc 1080  
gtcatggagg tcctcgccgc caagactggc tacgagaccg acatgatcga gtccgacatg 1140  
gagctcgaga ccgagctcgg cattgactcc atcaagcgtg tcgagattct ctccgagggt 1200  
caggccatgc tcaacgtcga ggccaaggac gtcgatgtc tcagccgcac tcgcactgtt 1260  
ggcgagggtc tcgatgccat gaaggctgag atcgccggca gctccgcccc ggcgcctgcc 1320  
gccgctgtc ctgctccggc tgctgccgt cctgcgcccc ctgccgtgc ccctgctgtc 1380  
tcgagcgagc ttctcgagaa ggccgagacc gtcgtcatgg aggtcctcgc cgccaagact 1440  
ggctacgaga ctgacatgat tgagtccgac atggagctcg agactgagct cggcattgac 1500  
tccatcaagc gtgtcgagat cctctccgag gttcaggcca tgctcaacgt cgaggccaag 1560  
gacgtcgatg ccctcagccg caccgcact gttggcgagg ttgtcgatgc catgaaggcc 1620  
gagatcgctg gtggctctgc ccgggcgcct gccgcgctg cccctgctcc ggctgccgcc 1680  
gccctgctg tctcgaacga gcttcttgag aaggccgaga ccgtcgtcat ggaggtcctc 1740  
gccgccaaaga ctggctacga gaccgacatg atcgagtccg acatggagct cgagaccgag 1800  
ctcggcattg actccatcaa gcgtgtcgag attctctccg aggttcaggc catgtcaac 1860  
gtcgaggcca aggacgtcga cgctctcagc cgcaactcga ctgttggcga ggtcgtcgat 1920  
gccaatgaagg ctgagatcgc tgggtggtct gccccggcgc ctgccgccgc tgctcctgcc 1980  
tcggctggcg ccgcgcctgc ggtcaagatt gactcgggtc acggcgctga ctgtgatgat 2040  
ctttccctga tgcacgccaa ggtggttgac atccgccgcc cggacgagct catcctggag 2100

cgccccgaga accgccccgt tctcgttgtc gatgacggca gcgagctcac cctcgccctg 2160  
 gtccgcgtcc tcggcgccctg cgccgttgtc ctgacctttg agggctctcca gctcgctcag 2220  
 cgcgctgggtg ccgctgccat ccgccacgtg ctcgccaaag atctttccgc ggagagcgcc 2280  
 gagaaggcca tcaaggagggc cgagcagcgc tttggcgctc tcggcggtt catctcgag 2340  
 caggcgaggc gcttcgagcc cgccgaaatc ctgggttca cgctcatgtg cgccaagtcc 2400  
 gccaaaggctt ccctctgcac ggctgtggct ggcgcccgcc cggcctttat cgggtgtggc 2460  
 cgcttgagc gccgcctcgg attcacttcg cagggcactt ctgacgcgct caagcgtgcc 2520  
 cagcgtgggtg ccattctttg cctctgcaag accatcggcc tcgagtggtc cgagtctgac 2580  
 gtcttttccc gcggcggtga cattgctcag ggcattgcacc ccgaggatgc cgccgtggcg 2640  
 attgtgcgag agatggcggtg cgctgacatt cgcattcgcg aggtcggcat tggcgcaaac 2700  
 cagcagcgct gcacgatccg tgcgcgaag ctcgagaccg gcaaccgcga gcgccagatc 2760  
 gccaaaggacg acgtgctgct cgtttctggc ggcgctcgcg gcattcacgcc tctttgcac 2820  
 cgggagatca cgcgccagat cgcgggcgcc aagtaattc tgcttgggcg cagcaaggtc 2880  
 tctgcgagcg aaccggcatg gtgcgctggc atcactgacg agaaggctgt gcaaaaggct 2940  
 gctaccagc agctcaagcg cgcctttagc gctggcgagg gcccgaagcc cagccccgc 3000  
 gctgtcacta agcttgtggg ctctgttctt ggcgctcgcg aggtgcgag ctctattgct 3060  
 gcgattgaag cgctcggcg caaggccatc tactcgtcgt gcgacgtgaa ctctgcgcc 3120  
 gacgtggcca aggcctgctg cgatgccgag tccagctcg gtgcccgcgt ctcgggcatc 3180  
 gttcatgcct cgggcgtgct ccgcgaccgt ctcatcgaga agaagctccc cgacgagttc 3240  
 gacgccgtct ttggcaccaa ggtcaccggt ctcgagaacc tctcgcgcg cgctcgaccgc 3300  
 gccaacctca agcacatggt cctcttcagc tcgctcgccg gcttccacgg caacgtcggc 3360  
 cagtctgact acgccatggc caacgaggcc cttacaaga tgggcctcga gctcgccaag 3420  
 gacgtctcgg tcaagtcgat ctgcttcggt ccctgggacg gtggcatggt gacgccgag 3480  
 ctcaagaagc agttccagga gatggcggtg cagatcatcc ccgcgaggg cgcgctgat 3540  
 accgtggcg gcacgtgct cggtcctcg ccggtgaga tccttgctcg caactggcg 3600  
 acccgtcca agaaggctcg ctcgacacc atcacctgc accgcaagat ttccgccaag 3660  
 tccaaccctt tctcagagga ccacgtcatc caggcgccgc gcgtgctgcc catgacgctg 3720  
 gccattggct cgctcgcgga gacctgcctc ggctcttcc ccggctactc gctctgggc 3780  
 attgacgacg ccagctctt caagggtgtc actgtcgacg gcgacgtcaa ctgcgagggtg 3840  
 accctcacc cgctcgacggc gccctcgggc cgctcaacg tccaggccac gctcaagacc 3900  
 ttttccagcg gcaagctggt ccgggcctac cgcgccgtca tcgtgctct caaccagggc 3960  
 gcgcccccg ccaacgccac catgcagccg ccctcgctcg atgccgatcc ggcgctccag 4020  
 ggctccgtct acgacggcaa gacctcttc cagggcccg ccttccgcg catcgatgac 4080  
 gtgctctcgt gcaccaagag ccagcttggt gccaaagtga gcgctgtccc cggtccgac 4140  
 gccgctcgc gcgagtttgc cagggacact gacgcccatt accccttctg gaacgacctg 4200  
 gcctttcagg ccattgctcgt ctgggtgcgc cgacgctcg gccaggctgc gctcccaac 4260  
 tcgatccagc gcacgtcca gcaccgccg gtccgcagg acaagccct ctacattacc 4320  
 ctccgtcca accagtcggg cgtcactcc cagcacaagc acgcccttca gttccacaac 4380  
 gagcaggcg atctcttcat tgatgtccag gcttcggtca tcgccacgga cagccttgcc 4440  
 ttctaa 4446

&lt;210&gt; 70

&lt;211&gt; 1481

&lt;212&gt; PRT

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 70.

Arg Cys Arg Arg Val Ser Pro Arg Arg Ala Ala Pro Pro Pro Pro Leu

1	5	10	15
Ala Arg Thr Pro Ala Arg Leu Ala Ala Pro Ala Val Ser Asn Glu Leu			
20	25	30	
Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr			
35	40	45	
Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu			
50	55	60	
Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln			
65	70	75	80
Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr			
85	90	95	
Arg Thr Val Gly Glu Val Val Asn Ala Met Lys Ala Glu Ile Ala Gly			
100	105	110	
Gly Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Gly Pro Ala Ala Ala			
115	120	125	
Ala Pro Ala Pro Ala Val Ser Ser Glu Leu Leu Glu Lys Ala Glu Thr			
130	135	140	
Val Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met			
145	150	155	160
Ile Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile			
165	170	175	
Lys Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu			
180	185	190	
Ala Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val			
195	200	205	
Val Asp Ala Met Lys Ala Glu Ile Ala Gly Ser Ser Ala Ser Ala Pro			
210	215	220	
Ala Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala			
225	230	235	240
Ala Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val			
245	250	255	
Val Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile			

260	265	270
Glu Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys		
275	280	285
Arg Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala		
290	295	300
Lys Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val		
305	310	315 320
Asp Ala Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala		
325	330	335
Ala Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Asn Glu		
340	345	350
Leu Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys		
355	360	365
Thr Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr		
370	375	380
Glu Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val		
385	390	395 400
Gln Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg		
405	410	415
Thr Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala		
420	425	430
Gly Ser Ser Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Pro Ala Ala		
435	440	445
Ala Ala Pro Ala Pro Ala Ala Ala Ala Pro Ala Val Ser Ser Glu Leu		
450	455	460
Leu Glu Lys Ala Glu Thr Val Val Met Glu Val Leu Ala Ala Lys Thr		
465	470	475 480
Gly Tyr Glu Thr Asp Met Ile Glu Ser Asp Met Glu Leu Glu Thr Glu		
485	490	495
Leu Gly Ile Asp Ser Ile Lys Arg Val Glu Ile Leu Ser Glu Val Gln		
500	505	510
Ala Met Leu Asn Val Glu Ala Lys Asp Val Asp Ala Leu Ser Arg Thr		

515	520	525
Arg Thr Val Gly Glu Val Val Asp Ala Met Lys Ala Glu Ile Ala Gly		
530	535	540
Gly Ser Ala Pro Ala Pro Ala Ala Ala Pro Ala Pro Ala Ala Ala		
545	550	555 560
Ala Pro Ala Val Ser Asn Glu Leu Leu Glu Lys Ala Glu Thr Val Val		
565	570	575
Met Glu Val Leu Ala Ala Lys Thr Gly Tyr Glu Thr Asp Met Ile Glu		
580	585	590
Ser Asp Met Glu Leu Glu Thr Glu Leu Gly Ile Asp Ser Ile Lys Arg		
595	600	605
Val Glu Ile Leu Ser Glu Val Gln Ala Met Leu Asn Val Glu Ala Lys		
610	615	620
Asp Val Asp Ala Leu Ser Arg Thr Arg Thr Val Gly Glu Val Val Asp		
625.	630	635 640
Ala Met Lys Ala Glu Ile Ala Gly Gly Ser Ala Pro Ala Pro Ala Ala		
645	650	655
Ala Ala Pro Ala Ser Ala Gly Ala Ala Pro Ala Val Lys Ile Asp Ser		
660	665	670
Val His Gly Ala Asp Cys Asp Asp Leu Ser Leu Met His Ala Lys Val		
675	680	685
Val Asp Ile Arg Arg Pro Asp Glu Leu Ile Leu Glu Arg Pro Glu Asn		
690	695	700
Arg Pro Val Leu Val Val Asp Asp Gly Ser Glu Leu Thr Leu Ala Leu		
705	710	715 720
Val Arg Val Leu Gly Ala Cys Ala Val Val Leu Thr Phe Glu Gly Leu		
725	730	735
Gln Leu Ala Gln Arg Ala Gly Ala Ala Ala Ile Arg His Val Leu Ala		
740	745	750
Lys Asp Leu Ser Ala Glu Ser Ala Glu Lys Ala Ile Lys Glu Ala Glu		
755	760	765
Gln Arg Phe Gly Ala Leu Gly Gly Phe Ile Ser Gln Gln Ala Glu Arg		



770	775	780
Phe Glu Pro Ala Glu Ile Leu Gly Phe Thr Leu Met Cys Ala Lys Phe		
785	790	795 800
Ala Lys Ala Ser Leu Cys Thr Ala Val Ala Gly Gly Arg Pro Ala Phe		
	805	810 815
Ile Gly Val Ala Arg Leu Asp Gly Arg Leu Gly Phe Thr Ser Gln Gly		
	820	825 830
Thr Ser Asp Ala Leu Lys Arg Ala Gln Arg Gly Ala Ile Phe Gly Leu		
	835	840 845
Cys Lys Thr Ile Gly Leu Glu Trp Ser Glu Ser Asp Val Phe Ser Arg		
	850	855 860
Gly Val Asp Ile Ala Gln Gly Met His Pro Glu Asp Ala Ala Val Ala		
865	870	875 880
Ile Val Arg Glu Met Ala Cys Ala Asp Ile Arg Ile Arg Glu Val Gly		
	885	890 895
Ile Gly Ala Asn Gln Gln Arg Cys Thr Ile Arg Ala Ala Lys Leu Glu		
	900	905 910
Thr Gly Asn Pro Gln Arg Gln Ile Ala Lys Asp Asp Val Leu Leu Val		
	915	920 925
Ser Gly Gly Ala Arg Gly Ile Thr Pro Leu Cys Ile Arg Glu Ile Thr		
	930	935 940
Arg Gln Ile Ala Gly Gly Lys Tyr Ile Leu Leu Gly Arg Ser Lys Val		
945	950	955 960
Ser Ala Ser Glu Pro Ala Trp Cys Ala Gly Ile Thr Asp Glu Lys Ala		
	965	970 975
Val Gln Lys Ala Ala Thr Gln Glu Leu Lys Arg Ala Phe Ser Ala Gly		
	980	985 990
Glu Gly Pro Lys Pro Thr Pro Arg Ala Val Thr Lys Leu Val Gly Ser		
	995	1000 1005
Val Leu Gly Ala Arg Glu Val Arg Ser Ser Ile Ala Ala Ile Glu Ala		
1010	1015	1020
Leu Gly Gly Lys Ala Ile Tyr Ser Ser Cys Asp Val Asn Ser Ala Ala		

WO 00/42195

1025	1030	1035	1040
Asp Val Ala Lys Ala Val Arg Asp Ala Glu Ser Gln Leu Gly Ala Arg			
	1045	1050	1055
Val Ser Gly Ile Val His Ala Ser Gly Val Leu Arg Asp Arg Leu Ile			
	1060	1065	1070
Glu Lys Lys Leu Pro Asp Glu Phe Asp Ala Val Phe Gly Thr Lys Val			
	1075	1080	1085
Thr Gly Leu Glu Asn Leu Leu Ala Ala Val Asp Arg Ala Asn Leu Lys			
	1090	1095	1100
His Met Val Leu Phe Ser Ser Leu Ala Gly Phe His Gly Asn Val Gly			
	1105	1110	1115
Gln Ser Asp Tyr Ala Met Ala Asn Glu Ala Leu Asn Lys Met Gly Leu			
	1125	1130	1135
Glu Leu Ala Lys Asp Val Ser Val Lys Ser Ile Cys Phe Gly Pro Trp			
	1140	1145	1150
Asp Gly Gly Met Val Thr Pro Gln Leu Lys Lys Gln Phe Gln Glu Met			
	1155	1160	1165
Gly Val Gln Ile Ile Pro Arg Glu Gly Gly Ala Asp Thr Val Ala Arg			
	1170	1175	1180
Ile Val Leu Gly Ser Ser Pro Ala Glu Ile Leu Val Gly Asn Trp Arg			
	1185	1190	1195
Thr Pro Ser Lys Lys Val Gly Ser Asp Thr Ile Thr Leu His Arg Lys			
	1205	1210	1215
Ile Ser Ala Lys Ser Asn Pro Phe Leu Glu Asp His Val Ile Gln Gly			
	1220	1225	1230
Arg Arg Val Leu Pro Met Thr Leu Ala Ile Gly Ser Leu Ala Glu Thr			
	1235	1240	1245
Cys Leu Gly Leu Phe Pro Gly Tyr Ser Leu Trp Ala Ile Asp Asp Ala			
	1250	1255	1260
Gln Leu Phe Lys Gly Val Thr Val Asp Gly Asp Val Asn Cys Glu Val			
	1265	1270	1275
Thr Leu Thr Pro Ser Thr Ala Pro Ser Gly Arg Val Asn Val Gln Ala			

	1285	1290	1295
Thr Leu Lys Thr Phe Ser Ser Gly Lys Leu Val Pro Ala Tyr Arg Ala	1300	1305	1310
Val Ile Val Leu Ser Asn Gln Gly Ala Pro Pro Ala Asn Ala Thr Met	1315	1320	1325
Gln Pro Pro Ser Leu Asp Ala Asp Pro Ala Leu Gln Gly Ser Val Tyr	1330	1335	1340
Asp Gly Lys Thr Leu Phe His Gly Pro Ala Phe Arg Gly Ile Asp Asp	1345	1350	1355
Val Leu Ser Cys Thr Lys Ser Gln Leu Val Ala Lys Cys Ser Ala Val	1365	1370	1375
Pro Gly Ser Asp Ala Ala Arg Gly Glu Phe Ala Thr Asp Thr Asp Ala	1380	1385	1390
His Asp Pro Phe Val Asn Asp Leu Ala Phe Gln Ala Met Leu Val Trp	1395	1400	1405
Val Arg Arg Thr Leu Gly Gln Ala Ala Leu Pro Asn Ser Ile Gln Arg	1410	1415	1420
Ile Val Gln His Arg Pro Val Pro Gln Asp Lys Pro Phe Tyr Ile Thr	1425	1430	1435
Leu Arg Ser Asn Gln Ser Gly Gly His Ser Gln His Lys His Ala Leu	1445	1450	1455
Gln Phe His Asn Glu Gln Gly Asp Leu Phe Ile Asp Val Gln Ala Ser	1460	1465	1470
Val Ile Ala Thr Asp Ser Leu Ala Phe	1475	1480	

&lt;210&gt; 71

&lt;211&gt; 5215

&lt;212&gt; DNA

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 71

tgccgtcttt gaggagcatg acccctccaa cgccgcctgc acggggccacg actccatttc 60  
 tgcgctctcg gcccgctgcg gcggtgaaag caacatgcgc atcgccatca ctggtatgga 120  
 cgccaccttt ggcgctctca agggactcga cgccttcgag cgcgccattt acaccggcgc 180

tcacgggtgcc atcccactcc cagaaaagcg ctggcgcttt ctcggaag acaaggactt 240  
 tcttgacctc tgcggcgctca aggccacccc gcaeggctgc tacattgaag atgttgaggt 300  
 cgacttccag cgcctccgca cggccatgac ccctgaagac atgctcctcc ctgagcagct 360  
 tctggccgctc accaccattg accgcgccat cctcgactcg ggaatgaaaa aggggtggcaa 420  
 tgtcgccgctc tttgtcggcc tcggcaccga cctcgagctc taccgtcacc gtgctcgcgt 480  
 cgctctcaag gagcgcgctc gccctgaagc ctccaagaag ctcaatgaca tgatgcagta 540  
 cattaacgac tgcggcacat ccacatcgta cacctcgtaac attggcaacc tcgtcgccac 600  
 gcgcgtctcg tcgcagtggt gcttcacggg cccctccttt acgatcaccg agggcaacaa 660  
 ctccgtctac cgctgcgcg agctcgga gtaacctctc gagaccggcg aggtcgatgg 720  
 cgctgctggt gcgggtgtcg atctctgagg cagtgcgaa aacctttacg tcaagtctcg 780  
 ccgcttcaag gtgtccacct ccgatacccc gcgcgcgagc tttgacgccg ccgccgatgg 840  
 ctactttgtc ggcgaggggt gcggtgcctt tgtgctcaag cgtgagacta gctgcaccaa 900  
 ggacgaccgt atctacgctt gcatggatgc catcgctcct ggcaacgtcc ctgagcgctg 960  
 cttgcgcgag gccctcgacc aggcgcgcgt caagccgggc gatatcgaga tgctcgagct 1020  
 cagcgccgac tccgcccgc acctcaagga cccgtccgtc ctgcccagg agctcactgc 1080  
 cgaggaggaa atcggcgggc ttcagacgat ccttcgtgac gatgacaagc tcccgcgcaa 1140  
 cgctcgaaacg ggcagtgtca aggccaccgt cggtagacac gggtatgcct ctggtgctgc 1200  
 cagcctcatc aaggtcgcg tttgcatcta caaccgctac ctgcccagca acggcgacga 1260  
 ctgggatgaa cccgcccctg aggcgcctcg ggacagcacc ctctttgctg gccagacctc 1320  
 gcgcgcttg ctcaagaacc ctggcgagcg tcgctatgct gccgtctcgg gcgtctccga 1380  
 gacgcgctcg tgctattccg tgctcctctc cgaagccgag ggccactacg agcgcgagaa 1440  
 ccgcatctcg ctcgacgagg aggcgcccga gctcattgtg cttcgcgcg actcccacga 1500  
 ggagatcctt ggtcgctcg acaagatccg cgagcgcttc ttgcagcca cgggcgcgc 1560  
 cccgcgcgag tccgagctca aggcgcaggc ccgcgcgcat ttcctcgagc tcctcggcga 1620  
 gacccttgcc caggatgccc cttcttcagg ctgcgaaaag cccctcgctc tcagcctcgt 1680  
 ctccacgccc tccaagctcc agcgcgaggt cgagctcgcg gccaaaggta tcccgcgctg 1740  
 cctcaagatg cgcgcgagtt ggagctcccc tgctggcagc cgctacgcgc ctgagccgct 1800  
 cgccagcgac cgcgtcgct tcatgtacgg cgaaggctgc agcccttact acggcatcac 1860  
 ccaagacatt caccgcattt ggcccgaact ccacgaggtc atcaacgaaa agacgaaccg 1920  
 tctctgggccc gaagcgagc gctgggtcat gccgcgcgcc agcttcaagt cggagctcga 1980  
 gagccagcag caagagtttg atcgcaacat gattgaaatg ttccgtcttg gaatcctcac 2040  
 ctcaattgcc ttaccaatc tggcgcgcca cgttctcaac atcacgcccga aggcgcctt 2100  
 tggcctcagt cttggcgaga tttccatgat ttttgcttt tccaagaaga acggtctcat 2160  
 ctccgaccag ctaccaagg atcttcgga gtcgcagctg tggacaagg ctctggcctg 2220  
 tgaatttaat gcgtgcgcg aggcctggg catccacag agtgteccca aggacgagtt 2280  
 ctggcaaggc tacattgtgc gcggcaccaa gcaggatata gaggcggcca tcgccccgga 2340  
 cagcaagtac gtgcgcctca ccatcatcaa tgatgccaac accgccctca ttagcgga 2400  
 gcccgacgcc tgcaaggctg cgatcgcgcg tctcggtggc aacattcctg cgcttcccgt 2460  
 gaccagggc atgtgcggcc actgcccga ggtgggacct tataccaagg atatcgccaa 2520  
 gatccatgcc aaccttgagt tcccgttgt cgacggcctt gacctctgga ccacaatcaa 2580  
 ccagaagcgc ctctgcccac gcgccacggg cgccaaggac gaatgggccc cttcttctt 2640  
 tggcgagtac gccggccagc tctacgagaa gcaggctaac tcccccaaa tcgtcgagac 2700  
 cattacaag caaaactacg acgtctttgt cgaggttggg cccaacaacc accgtagcac 2760  
 cgagtgccg accacgcttg gtccccagc caaccacctt gctggcgcca tcgacaagca 2820  
 gaacgaggat gcttgagca ccatcgtaa gcttgtggct tcgctcaagg cccacctgt 2880  
 tcttgccgctc acgatctcgc cgctgtacca ctccaagctt gtggcgagg ctcaggcttg 2940  
 ctacgtcgcg ctctgcaagg gtgaaaagcc caagaagaac aagtttgtgc gcaagattca 3000  
 gctcaacggc cgcttcaaca gcaaggcgga ccccatctcc tcggccgac ttgccagctt 3060

```

tccgcctgcg gaccctgcc a ttgaagccgc catctcgagc cgcattcatga agcctgtcgc 3120
tcccaagtgc tacgcgcgtc tcaacattga cgagcaggac gagacccgag atccgatcct 3180
caacaaggac aacgcgcgtt cttcttcttc ttcttcttct tcttcttctt cttcttcttc 3240
ttctccgtcg cctgtctcctt cggcccccgt gcaaaagaag gctgctcccg ccgcggagac 3300
caaggctggt gcttcggctg acgcacttcg cagtgccttg ctcgatctcg acagtatgct 3360
tgcgtgagc tctgccagtg cctccggcaa ccttggtgag actgcgccta gcgacgcctc 3420
ggtcattgtg ccgccctgca acattgcgga tctcggcagc cgcgccttca tgaaaacgta 3480
cgggtgttcg gcgcctctgt acacgggcgc catggccaag ggcattgcct ctgcggacct 3540
cgtcattgcc gccggccgcc agggcatcct tgcgtccttt ggcgcggcgc gacttcccat 3600
gcaggttggt cgtgagtgca tcgaaaagat tcaggccgcc ctgccaatg gcccgtagc 3660
tgtcaacctt atccattctc cctttgacag caacctcgaa aagggaatg tcgatctctt 3720
cctcgagaag ggtgtcacct ttgtcgaggc ctccgccttt atgacgtca ccccgaggt 3780
cgtgcgttac cgcgcggctg gcctcacgcg caacgccgac ggctcggta acatccgcaa 3840
ccgtatcatt ggcaaggctt cgcgcaccga gctcgcgag atgttcatgc gtectgcgc 3900
cgagcacctt cttcagaagc tcattgtctt cggcgagatc aaccaggagc aggccgagct 3960
cgcccgccgt gttcccgtcg ctgacgacat cgcggtcgaa gctgactcgg gtggccacac 4020
cgacaaccgc cccatccacg tcattctgcc cctcatcacc aaccttcgcg accgccttca 4080
ccgcgagtgc ggctacccgg ccaaccttcg cgtccgtgtg ggcgcggcgc gtggcattgg 4140
gtgccccag gcggcgctgg ccaccttcaa catgggtgcc tcctttattg tcaccggcac 4200
cgtgaaccag gtcgccaagc agtcgggcac gtgcgacaat gtgcgcaagc agctcgcgaa 4260
ggccacttac tcggacgtat gcatggcccc ggctgccgac atgttcgagg aaggcgtcaa 4320
gcttcaggtc ctcaagaagg gaaccatgtt tccctcgcgc gccacaagc tctacgagct 4380
cttttgcaag tacgactcgt tcgagtcctt gcccccgca gagcttgcgc gcgtcgagaa 4440
gcgcattctt agccgcgcgc tcgaagaggt ctgggacgag accaaaaact ttacattaa 4500
ccgtcttcac aaccggaga agatccagcg cgcgcgagc gacccaagc tcaagatgtc 4560
gctgtgcttt cgctggtacc tgagcctggc gagccgctgg gccaacactg gagcttccga 4620
tcgcgtcatg gactaccagg tctggtgcgg tcctgccatt ggttccttca acgatttcat 4680
caagggaact taccttgatc cggccgtcgc aaacgagta ccgtgcgtcg ttcagattaa 4740
caagcagatc cttcgtggag cgtgcttctt gcgccgtctc gaaattctgc gcaacgcacg 4800
cctttccgat ggcgctgccg ctctgtggc cagcatcgat gacacatacg tcccgccga 4860
gaagctgtaa gtaagctctc atatatgta gttgcgtgag accgacacga agataatatc 4920
acatacgctt ttgtttgttc tttcaattat ttgtctgtgc ttcattgtgc tcctcagtat 4980
ctagctggcg gctcttatct tcttttaaaa tatctggaca aggacaaaaa caagaataaa 5040
ggcgagaaga tgtgaatttc atttcgactt gagaactcga agagcattga tgcggttagt 5100
atatgggtat tttccagaca cttttcatca tcatcatcat catcatcatt atgaagaagt 5160
agtagctgat aaagtagact cactgtttgc agcgagaaaa aaaaaaaaaa aaaaa 5215

```

&lt;210&gt; 72

&lt;211&gt; 1622

&lt;212&gt; PRT

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 72

Ala Val Phe Glu Glu His Asp Pro Ser Asn Ala Ala Cys Thr Gly His

1

5

10

15

Asp Ser Ile Ser Ala Leu Ser Ala Arg Cys Gly Gly Glu Ser Asn Met

20

25

30

Arg Ile Ala Ile Thr Gly Met Asp Ala Thr Phe Gly Ala Leu Lys Gly  
 35 40 45

Leu Asp Ala Phe Glu Arg Ala Ile Tyr Thr Gly Ala His Gly Ala Ile  
 50 55 60

Pro Leu Pro Glu Lys Arg Trp Arg Phe Leu Gly Lys Asp Lys Asp Phe  
 65 70 75 80

Leu Asp Leu Cys Gly Val Lys Ala Thr Pro His Gly Cys Tyr Ile Glu  
 85 90 95

Asp Val Glu Val Asp Phe Gln Arg Leu Arg Thr Pro Met Thr Pro Glu  
 100 105 110

Asp Met Leu Leu Pro Gln Gln Leu Leu Ala Val Thr Thr Ile Asp Arg  
 115 120 125

Ala Ile Leu Asp Ser Gly Met Lys Lys Gly Gly Asn Val Ala Val Phe  
 130 135 140

Val Gly Leu Gly Thr Asp Leu Glu Leu Tyr Arg His Arg Ala Arg Val  
 145 150 155 160

Ala Leu Lys Glu Arg Val Arg Pro Glu Ala Ser Lys Lys Leu Asn Asp  
 165 170 175

Met Met Gln Tyr Ile Asn Asp Cys Gly Thr Ser Thr Ser Tyr Thr Ser  
 180 185 190

Tyr Ile Gly Asn Leu Val Ala Thr Arg Val Ser Ser Gln Trp Gly Phe  
 195 200 205

Thr Gly Pro Ser Phe Thr Ile Thr Glu Gly Asn Asn Ser Val Tyr Arg  
 210 215 220

Cys Ala Glu Leu Gly Lys Tyr Leu Leu Glu Thr Gly Glu Val Asp Gly  
 225 230 235 240

Val Val Val Ala Gly Val Asp Leu Cys Gly Ser Ala Glu Asn Leu Tyr  
 245 250 255

Val Lys Ser Arg Arg Phe Lys Val Ser Thr Ser Asp Thr Pro Arg Ala  
 260 265 270

Ser Phe Asp Ala Ala Ala Asp Gly Tyr Phe Val Gly Glu Gly Cys Gly  
 275 280 285

Ala Phe Val Leu Lys Arg Glu Thr Ser Cys Thr Lys Asp Asp Arg Ile  
 290 295 300  
 Tyr Ala Cys Met Asp Ala Ile Val Pro Gly Asn Val Pro Ser Ala Cys  
 305 310 315 320  
 Leu Arg Glu Ala Leu Asp Gln Ala Arg Val Lys Pro Gly Asp Ile Glu  
 325 330 335  
 Met Leu Glu Leu Ser Ala Asp Ser Ala Arg His Leu Lys Asp Pro Ser  
 340 345 350  
 Val Leu Pro Lys Glu Leu Thr Ala Glu Glu Glu Ile Gly Gly Leu Gln  
 355 360 365  
 Thr Ile Leu Arg Asp Asp Asp Lys Leu Pro Arg Asn Val Ala Thr Gly  
 370 375 380  
 Ser Val Lys Ala Thr Val Gly Asp Thr Gly Tyr Ala Ser Gly Ala Ala  
 385 390 395 400  
 Ser Leu Ile Lys Ala Ala Leu Cys Ile Tyr Asn Arg Tyr Leu Pro Ser  
 405 410 415  
 Asn Gly Asp Asp Trp Asp Glu Pro Ala Pro Glu Ala Pro Trp Asp Ser  
 420 425 430  
 Thr Leu Phe Ala Cys Gln Thr Ser Arg Ala Trp Leu Lys Asn Pro Gly  
 435 440 445  
 Glu Arg Arg Tyr Ala Ala Val Ser Gly Val Ser Glu Thr Arg Ser Cys  
 450 455 460  
 Tyr Ser Val Leu Leu Ser Glu Ala Glu Gly His Tyr Glu Arg Glu Asn  
 465 470 475 480  
 Arg Ile Ser Leu Asp Glu Glu Ala Pro Lys Leu Ile Val Leu Arg Ala  
 485 490 495  
 Asp Ser His Glu Glu Ile Leu Gly Arg Leu Asp Lys Ile Arg Glu Arg  
 500 505 510  
 Phe Leu Gln Pro Thr Gly Ala Ala Pro Arg Glu Ser Glu Leu Lys Ala  
 515 520 525  
 Gln Ala Arg Arg Ile Phe Leu Glu Leu Leu Gly Glu Thr Leu Ala Gln  
 530 535 540

Asp Ala Ala Ser Ser Gly Ser Gln Lys Pro Leu Ala Leu Ser Leu Val  
 545 550 555 560

Ser Thr Pro Ser Lys Leu Gln Arg Glu Val Glu Leu Ala Ala Lys Gly  
 565 570 575

Ile Pro Arg Cys Leu Lys Met Arg Arg Asp Trp Ser Ser Pro Ala Gly  
 580 585 590

Ser Arg Tyr Ala Pro Glu Pro Leu Ala Ser Asp Arg Val Ala Phe Met  
 595 600 605

Tyr Gly Glu Gly Arg Ser Pro Tyr Tyr Gly Ile Thr Gln Asp Ile His  
 610 615 620

Arg Ile Trp Pro Glu Leu His Glu Val Ile Asn Glu Lys Thr Asn Arg  
 625 630 635 640

Leu Trp Ala Glu Gly Asp Arg Trp Val Met Pro Arg Ala Ser Phe Lys  
 645 650 655

Ser Glu Leu Glu Ser Gln Gln Gln Glu Phe Asp Arg Asn Met Ile Glu  
 660 665 670

Met Phe Arg Leu Gly Ile Leu Thr Ser Ile Ala Phe Thr Asn Leu Ala  
 675 680 685

Arg Asp Val Leu Asn Ile Thr Pro Lys Ala Ala Phe Gly Leu Ser Leu  
 690 695 700

Gly Glu Ile Ser Met Ile Phe Ala Phe Ser Lys Lys Asn Gly Leu Ile  
 705 710 715 720

Ser Asp Gln Leu Thr Lys Asp Leu Arg Glu Ser Asp Val Trp Asn Lys  
 725 730 735

Ala Leu Ala Val Glu Phe Asn Ala Leu Arg Glu Ala Trp Gly Ile Pro  
 740 745 750

Gln Ser Val Pro Lys Asp Glu Phe Trp Gln Gly Tyr Ile Val Arg Gly  
 755 760 765

Thr Lys Gln Asp Ile Glu Ala Ala Ile Ala Pro Asp Ser Lys Tyr Val  
 770 775 780

Arg Leu Thr Ile Ile Asn Asp Ala Asn Thr Ala Leu Ile Ser Gly Lys  
 785 790 795 800



Pro Asp Ala Cys Lys Ala Ala Ile Ala Arg Leu Gly Gly Asn Ile Pro  
                     805                    810                    815

Ala Leu Pro Val Thr Gln Gly Met Cys Gly His Cys Pro Glu Val Gly  
                     820                    825                    830

Pro Tyr Thr Lys Asp Ile Ala Lys Ile His Ala Asn Leu Glu Phe Pro  
                     835                    840                    845

Val Val Asp Gly Leu Asp Leu Trp Thr Thr Ile Asn Gln Lys Arg Leu  
                     850                    855                    860

Val Pro Arg Ala Thr Gly Ala Lys Asp Glu Trp Ala Pro Ser Ser Phe  
                     865                    870                    875                    880

Gly Glu Tyr Ala Gly Gln Leu Tyr Glu Lys Gln Ala Asn Phe Pro Gln  
                     885                    890                    895

Ile Val Glu Thr Ile Tyr Lys Gln Asn Tyr Asp Val Phe Val Glu Val  
                     900                    905                    910

Gly Pro Asn Asn His Arg Ser Thr Ala Val Arg Thr Thr Leu Gly Pro  
                     915                    920                    925

Gln Arg Asn His Leu Ala Gly Ala Ile Asp Lys Gln Asn Glu Asp Ala  
                     930                    935                    940

Trp Thr Thr Ile Val Lys Leu Val Ala Ser Leu Lys Ala His Leu Val  
                     945                    950                    955                    960

Pro Gly Val Thr Ile Ser Pro Leu Tyr His Ser Lys Leu Val Ala Glu  
                     965                    970                    975

Ala Gln Ala Cys Tyr Ala Ala Leu Cys Lys Gly Glu Lys Pro Lys Lys  
                     980                    985                    990

Asn Lys Phe Val Arg Lys Ile Gln Leu Asn Gly Arg Phe Asn Ser Lys  
                     995                    1000                    1005

Ala Asp Pro Ile Ser Ser Ala Asp Leu Ala Ser Phe Pro Pro Ala Asp  
                     1010                    1015                    1020

Pro Ala Ile Glu Ala Ala Ile Ser Ser Arg Ile Met Lys Pro Val Ala  
                     1025                    1030                    1035                    1040

Pro Lys Phe Tyr Ala Arg Leu Asn Ile Asp Glu Gln Asp Glu Thr Arg  
                     1045                    1050                    1055

Asp Pro Ile Leu Asn Lys Asp Asn Ala Pro Ser Ser Ser Ser Ser Ser  
 1060 1065 1070

Ser Ser Ser Ser Ser Ser Ser Ser Ser Pro Ser Pro Ala Pro Ser Ala  
 1075 1080 1085

Pro Val Gln Lys Lys Ala Ala Pro Ala Ala Glu Thr Lys Ala Val Ala  
 1090 1095 1100

Ser Ala Asp Ala Leu Arg Ser Ala Leu Leu Asp Leu Asp Ser Met Leu  
 1105 1110 1115 1120

Ala Leu Ser Ser Ala Ser Ala Ser Gly Asn Leu Val Glu Thr Ala Pro  
 1125 1130 1135

Ser Asp Ala Ser Val Ile Val Pro Pro Cys Asn Ile Ala Asp Leu Gly  
 1140 1145 1150

Ser Arg Ala Phe Met Lys Thr Tyr Gly Val Ser Ala Pro Leu Tyr Thr  
 1155 1160 1165

Gly Ala Met Ala Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala  
 1170 1175 1180

Gly Arg Gln Gly Ile Leu Ala Ser Phe Gly Ala Gly Gly Leu Pro Met  
 1185 1190 1195 1200

Gln Val Val Arg Glu Ser Ile Glu Lys Ile Gln Ala Ala Leu Pro Asn  
 1205 1210 1215

Gly Pro Tyr Ala Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn Leu  
 1220 1225 1230

Glu Lys Gly Asn Val Asp Leu Phe Leu Glu Lys Gly Val Thr Phe Val  
 1235 1240 1245

Glu Ala Ser Ala Phe Met Thr Leu Thr Pro Gln Val Val Arg Tyr Arg  
 1250 1255 1260

Ala Ala Gly Leu Thr Arg Asn Ala Asp Gly Ser Val Asn Ile Arg Asn  
 1265 1270 1275 1280

Arg Ile Ile Gly Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe Met  
 1285 1290 1295

Arg Pro Ala Pro Glu His Leu Leu Gln Lys Leu Ile Ala Ser Gly Glu  
 1300 1305 1310

Ile Asn Gln Glu Gln Ala Glu Leu Ala Arg Arg Val Pro Val Ala Asp  
 1315 1320 1325

Asp Ile Ala Val Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg Pro  
 1330 1335 1340

Ile His Val Ile Leu Pro Leu Ile Ile Asn Leu Arg Asp Arg Leu His  
 1345 1350 1355 1360

Arg Glu Cys Gly Tyr Pro Ala Asn Leu Arg Val Arg Val Gly Ala Gly  
 1365 1370 1375

Gly Gly Ile Gly Cys Pro Gln Ala Ala Leu Ala Thr Phe Asn Met Gly  
 1380 1385 1390

Ala Ser Phe Ile Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser  
 1395 1400 1405

Gly Thr Cys Asp Asn Val Arg Lys Gln Leu Ala Lys Ala Thr Tyr Ser  
 1410 1415 1420

Asp Val Cys Met Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val Lys  
 1425 1430 1435 1440

Leu Gln Val Leu Lys Lys Gly Thr Met Phe Pro Ser Arg Ala Asn Lys  
 1445 1450 1455

Leu Tyr Glu Leu Phe Cys Lys Tyr Asp Ser Phe Glu Ser Met Pro Pro  
 1460 1465 1470

Ala Glu Leu Ala Arg Val Glu Lys Arg Ile Phe Ser Arg Ala Leu Glu  
 1475 1480 1485

Glu Val Trp Asp Glu Thr Lys Asn Phe Tyr Ile Asn Arg Leu His Asn  
 1490 1495 1500

Pro Glu Lys Ile Gln Arg Ala Glu Arg Asp Pro Lys Leu Lys Met Ser  
 1505 1510 1515 1520

Leu Cys Phe Arg Trp Tyr Leu Ser Leu Ala Ser Arg Trp Ala Asn Thr  
 1525 1530 1535

Gly Ala Ser Asp Arg Val Met Asp Tyr Gln Val Trp Cys Gly Pro Ala  
 1540 1545 1550

Ile Gly Ser Phe Asn Asp Phe Ile Lys Gly Thr Tyr Leu Asp Pro Ala  
 1555 1560 1565

Val Ala Asn Glu Tyr Pro Cys Val Val Gln Ile Asn Lys Gln Ile Leu  
 1570 1575 1580

Arg Gly Ala Cys Phe Leu Arg Arg Leu Glu Ile Leu Arg Asn Ala Arg  
 1585 1590 1595 1600

Leu Ser Asp Gly Ala Ala Ala Leu Val Ala Ser Ile Asp Asp Thr Tyr  
 1605 1610 1615

Val Pro Ala Glu Lys Leu  
 1620

<210> 73

<211> 1551

<212> PRT

<213> Schizochytrium aggregatum

<400> 73

Arg Ala Glu Ala Gly Arg Glu Pro Glu Pro Ala Pro Gln Ile Thr Ser  
 1 5 10 15

Thr Ala Ala Glu Ser Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln  
 20 25 30

Gln Gln Gln Gln Pro Arg Glu Gly Asp Lys Glu Lys Ala Ala Glu Thr  
 35 40 45

Met Ala Leu Arg Val Lys Thr Asn Lys Lys Pro Cys Trp Glu Met Thr  
 50 55 60

Lys Glu Glu Leu Thr Ser Gly Lys Thr Glu Val Phe Asn Tyr Glu Glu  
 65 70 75 80

Leu Leu Glu Phe Ala Glu Gly Asp Ile Ala Lys Val Phe Gly Pro Glu  
 85 90 95

Phe Ala Val Ile Asp Lys Tyr Pro Arg Arg Val Arg Leu Pro Ala Arg  
 100 105 110

Glu Tyr Leu Leu Val Thr Arg Val Thr Leu Met Asp Ala Glu Val Asn  
 115 120 125

Asn Tyr Arg Val Gly Ala Arg Met Val Thr Glu Tyr Asp Leu Pro Val  
 130 135 140

Asn Gly Glu Leu Ser Glu Gly Gly Asp Cys Pro Trp Ala Val Leu Val

145		150		155		160
Glu Ser Gly Gln Cys Asp Leu Met Leu Ile Ser Tyr Met Gly Ile Asp						
		165		170		175
Phe Gln Asn Gln Gly Asp Arg Val Tyr Arg Leu Leu Asn Thr Thr Leu						
		180		185		190
Thr Phe Tyr Gly Val Ala His Glu Gly Glu Thr Leu Glu Tyr Asp Ile						
		195		200		205
Arg Val Thr Gly Phe Ala Lys Arg Leu Asp Gly Gly Ile Ser Met Phe						
		210		215		220
Phe Phe Glu Tyr Asp Cys Tyr Val Asn Gly Arg Leu Leu Ile Glu Met						
225		230		235		240
Arg Asp Gly Cys Ala Gly Phe Phe Thr Asn Glu Glu Leu Asp Ala Gly						
		245		250		255
Lys Gly Val Val Phe Thr Arg Gly Asp Leu Ala Ala Arg Ala Lys Ile						
		260		265		270
Pro Lys Gln Asp Val Ser Pro Tyr Ala Val Ala Pro Cys Leu His Lys						
		275		280		285
Thr Lys Leu Asn Glu Lys Glu Met Gln Thr Leu Val Asp Lys Asp Trp						
		290		295		300
Ala Ser Val Phe Gly Ser Lys Asn Gly Met Pro Glu Ile Asn Tyr Lys						
305		310		315		320
Leu Cys Ala Arg Lys Met Leu Met Ile Asp Arg Val Thr Ser Ile Asp						
		325		330		335
His Lys Gly Gly Val Tyr Gly Leu Gly Gln Leu Val Gly Glu Lys Ile						
		340		345		350
Leu Glu Arg Asp His Trp Tyr Phe Pro Cys His Phe Val Lys Asp Gln						
		355		360		365
Val Met Ala Gly Ser Leu Val Ser Asp Gly Cys Ser Gln Met Leu Lys						
		370		375		380
Met Tyr Met Ile Trp Leu Gly Leu His Leu Thr Thr Gly Pro Phe Asp						
385		390		395		400
Phe Arg Pro Val Asn Gly His Pro Asn Lys Val Arg Cys Arg Gly Gln						

405	410	415
Ile Ser Pro His Lys Gly Lys Leu Val Tyr Val Met Glu Ile Lys Glu		
420	425	430
Met Gly Phe Asp Glu Asp Asn Asp Pro Tyr Ala Ile Ala Asp Val Asn		
435	440	445
Ile Ile Asp Val Asp Phe Glu Lys Gly Gln Asp Phe Ser Leu Asp Arg		
450	455	460
Ile Ser Asp Tyr Gly Lys Gly Asp Leu Asn Lys Lys Ile Val Val Asp		
465	470	475
Phe Lys Gly Ile Ala Leu Lys Met Gln Lys Arg Ser Thr Asn Lys Asn		
485	490	495
Pro Ser Lys Val Gln Pro Val Phe Ala Asn Gly Ala Ala Thr Val Gly		
500	505	510
Pro Glu Ala Ser Lys Ala Ser Ser Gly Ala Ser Ala Ser Ala Ser Ala		
515	520	525
Ala Pro Ala Lys Pro Ala Phe Ser Ala Asp Val Leu Ala Pro Lys Pro		
530	535	540
Val Ala Leu Pro Glu His Ile Leu Lys Gly Asp Ala Leu Ala Pro Lys		
545	550	555
Glu Met Ser Trp His Pro Met Ala Arg Ile Pro Gly Asn Pro Thr Pro		
565	570	575
Ser Phe Ala Pro Ser Ala Tyr Lys Pro Arg Asn Ile Ala Phe Thr Pro		
580	585	590
Phe Pro Gly Asn Pro Asn Asp Asn Asp His Thr Pro Gly Lys Met Pro		
595	600	605
Leu Thr Trp Phe Asn Met Ala Glu Phe Met Ala Gly Lys Val Ser Met		
610	615	620
Cys Leu Gly Pro Glu Phe Ala Lys Phe Asp Asp Ser Asn Thr Ser Arg		
625	630	635
Ser Pro Ala Trp Asp Leu Ala Leu Val Thr Arg Ala Val Ser Val Ser		
645	650	655
Asp Leu Lys His Val Asn Tyr Arg Asn Ile Asp Leu Asp Pro Ser Lys		

660	665	670
Gly Thr Met Val Gly Glu Phe Asp Cys Pro Ala Asp Ala Trp Phe Tyr		
675	680	685
Lys Gly Ala Cys Asn Asp Ala His Met Pro Tyr Ser Ile Leu Met Glu		
690	695	700
Ile Ala Leu Gln Thr Ser Gly Val Leu Thr Ser Val Leu Lys Ala Pro		
705	710	715
Leu Thr Met Glu Lys Asp Asp Ile Leu Phe Arg Asn Leu Asp Ala Asn		
725	730	735
Ala Glu Phe Val Arg Ala Asp Leu Asp Tyr Arg Gly Lys Thr Ile Arg		
740	745	750
Asn Val Thr Lys Cys Thr Gly Tyr Ser Met Leu Gly Glu Met Gly Val		
755	760	765
His Arg Phe Thr Phe Glu Leu Tyr Val Asp Asp Val Leu Phe Tyr Lys		
770	775	780
Gly Ser Thr Ser Phe Gly Trp Phe Val Pro Glu Val Phe Ala Ala Gln		
785	790	795
Ala Gly Leu Asp Asn Gly Arg Lys Ser Glu Pro Trp Phe Ile Glu Asn		
805	810	815
Lys Val Pro Ala Ser Gln Val Ser Ser Phe Asp Val Arg Pro Asn Gly		
820	825	830
Ser Gly Arg Thr Ala Ile Phe Ala Asn Ala Pro Ser Gly Ala Gln Leu		
835	840	845
Asn Arg Arg Thr Asp Gln Gly Gln Tyr Leu Asp Ala Val Asp Ile Val		
850	855	860
Ser Gly Ser Gly Lys Lys Ser Leu Gly Tyr Ala His Gly Ser Lys Thr		
865	870	875
Val Asn Pro Asn Asp Trp Phe Phe Ser Cys His Phe Trp Phe Asp Ser		
885	890	895
Val Met Pro Gly Ser Leu Gly Val Glu Ser Met Phe Gln Leu Val Glu		
900	905	910
Ala Ile Ala Ala His Glu Asp Leu Ala Gly Lys Ala Arg His Cys Gln		

915	920	925
Pro His Leu Cys Ala Arg	Pro Arg Ala Arg Ser Ser Trp Lys Tyr Arg	
930	935	940
Gly Gln Leu Thr Pro Lys Ser Lys Lys Met Asp Ser Glu Val His Ile		
945	950	955
Val Ser Val Asp Ala His Asp Gly Val Val Asp Leu Val Ala Asp Gly		
965	970	975
Phe Leu Trp Ala Asp Ser Leu Arg Val Tyr Ser Val Ser Asn Ile Arg		
980	985	990
Val Arg Ile Ala Ser Gly Glu Ala Pro Ala Ala Ala Ser Ser Ala Ala		
995	1000	1005
Ser Val Gly Ser Ser Ala Ser Ser Val Glu Arg Thr Arg Ser Ser Pro		
1010	1015	1020
Ala Val Ala Ser Gly Pro Ala Gln Thr Ile Asp Leu Lys Gln Leu Lys		
1025	1030	1035
Thr Glu Leu Leu Glu Leu Asp Ala Pro Leu Tyr Leu Ser Gln Asp Pro		
1045	1050	1055
Thr Ser Gly Gln Leu Lys Lys His Thr Asp Val Ala Ser Gly Gln Ala		
1060	1065	1070
Thr Ile Val Gln Pro Cys Thr Leu Gly Asp Leu Gly Asp Arg Ser Phe		
1075	1080	1085
Met Glu Thr Tyr Gly Val Val Ala Pro Leu Tyr Thr Gly Ala Met Ala		
1090	1095	1100
Lys Gly Ile Ala Ser Ala Asp Leu Val Ile Ala Ala Gly Lys Arg Lys		
1105	1110	1115
Ile Leu Gly Ser Phe Gly Ala Gly Gly Leu Pro Met His His Val Arg		
1125	1130	1135
Ala Ala Leu Glu Lys Ile Gln Ala Ala Leu Pro Gln Gly Pro Tyr Ala		
1140	1145	1150
Val Asn Leu Ile His Ser Pro Phe Asp Ser Asn Leu Glu Lys Gly Asn		
1155	1160	1165
Val Asp Leu Phe Leu Glu Lys Gly Val Thr Val Val Glu Ala Ser Ala		



1170	1175	1180
Phe Met Thr Leu Thr Pro Gln Val Val Arg Tyr Arg Ala Ala Gly Leu		
1185	1190	1195 1200
Ser Arg Asn Ala Asp Gly Ser Val Asn Ile Arg Asn Arg Ile Ile Gly		
1205	1210	1215
Lys Val Ser Arg Thr Glu Leu Ala Glu Met Phe Ile Arg Pro Ala Pro		
1220	1225	1230
Glu His Leu Leu Glu Lys Leu Ile Ala Ser Gly Glu Ile Thr Gln Glu		
1235	1240	1245
Gln Ala Glu Leu Ala Arg Arg Val Pro Val Ala Asp Asp Ile Ala Val		
1250	1255	1260
Glu Ala Asp Ser Gly Gly His Thr Asp Asn Arg Pro Ile His Val Ile		
1265	1270	1275 1280
Leu Pro Leu Ile Ile Asn Leu Arg Asn Arg Leu His Arg Glu Cys Gly		
1285	1290	1295
Tyr Pro Ala His Leu Arg Val Arg Val Gly Ala Gly Gly Gly Val Gly		
1300	1305	1310
Cys Pro Gln Ala Ala Ala Ala Ala Leu Thr Met Gly Ala Ala Phe Ile		
1315	1320	1325
Val Thr Gly Thr Val Asn Gln Val Ala Lys Gln Ser Gly Thr Cys Asp		
1330	1335	1340
Asn Val Arg Lys Gln Leu Ser Gln Ala Thr Tyr Ser Asp Ile Cys Met		
1345	1350	1355 1360
Ala Pro Ala Ala Asp Met Phe Glu Glu Gly Val Lys Leu Gln Val Leu		
1365	1370	1375
Lys Lys Gly Thr Met Phe Pro Ser Arg Ala Asn Lys Leu Tyr Glu Leu		
1380	1385	1390
Phe Cys Lys Tyr Asp Ser Phe Asp Ser Met Pro Pro Ala Glu Leu Glu		
1395	1400	1405
Arg Ile Glu Lys Arg Ile Phe Lys Arg Ala Leu Gln Glu Val Trp Glu		
1410	1415	1420
Glu Thr Lys Asp Phe Tyr Ile Asn Gly Leu Lys Asn Pro Glu Lys Ile		

1425	1430	1435	1440
Gln Arg Ala Glu His Asp Pro Lys Leu Lys Met Ser Leu Cys Phe Arg			
1445	1450	1455	
Trp Tyr Leu Gly Leu Ala Ser Arg Trp Ala Asn Met Gly Ala Pro Asp			
1460	1465	1470	
Arg Val Met Asp Tyr Gln Val Trp Cys Gly Pro Ala Ile Gly Ala Phe			
1475	1480	1485	
Asn Asp Phe Ile Lys Gly Thr Tyr Leu Asp Pro Ala Val Ser Asn Glu			
1490	1495	1500	
Tyr Pro Cys Val Val Gln Ile Asn Leu Gln Ile Leu Arg Gly Ala Cys			
1505	1510	1515	1520
Tyr Leu Arg Arg Leu Asn Ala Leu Arg Asn Asp Pro Arg Ile Asp Leu			
1525	1530	1535	
Glu Thr Glu Asp Ala Ala Phe Val Tyr Glu Pro Thr Asn Ala Leu			
1540	1545	1550	

&lt;210&gt; 74

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 74

taccgcggca agactatccg caacgtcacc

30

&lt;210&gt; 75

&lt;211&gt; 30

&lt;212&gt; DNA

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 75

gccgtcgtgg gcgtccacgg acacgatgtg

30

&lt;210&gt; 76

&lt;211&gt; 4767

&lt;212&gt; DNA

&lt;213&gt; Schizochytrium aggregatum

&lt;400&gt; 76

cgagcagagg ccggccgcga gcccagagccc gcgccgcaga tcactagtag cgctgcggaa 60  
tcacagcagc agcagcagca gcagcagcag cagcagcagc agcagcagcc acgagagggga 120

gataaagaaa aagcggcaga gacgatggcg ctccgtgtca agacgaacaa gaagccatgc 180  
tgggagatga ccaaggagga gctgaccagc ggcaagaccg aggtgttcaa ctatgaggaa 240  
ctcctcgagt tcgagaggg cgacatcgcc aaggtcttcg gacccgagtt cgccgtcatc 300  
gacaagtacc cgcgcgcgt gcgcctgccc gcccgcgagt acctgtcgt gacccgcgtc 360  
accctcatgg acgccgaggt caacaactac cgcgtcggcg cccgcatggt caccgagtac 420  
gatctccccg tcaacggaga gctctccgag ggcggagact gcccctgggc cgtcctggtc 480  
gagagtggcc agtgcgatct catgctcatc tcctacatgg gcattgactt ccagaaccag 540  
ggcgaccgag tctaccgcct gctcaacacc acgctcacct tttaaggcgt ggcccacgag 600  
ggcgagaccc tcgagtacga cattcgcgtc accggttcg ccaagcgtct cgacggcggc 660  
atctccatgt tcttcttcga gtacgactgc tacgtcaacg gccgcctcct catcgagatg 720  
cgcgatgggt gcgcgggctt cttcaccaac gaggagctcg acgccggcaa gggcgtcgtc 780  
ttcacccgag gcgacctcgc cgcgcgcgc aagatcccaa agcaggacgt cccccctac 840  
gccgtcgcgc cctgcctcca caagaccaag ctcaacgaaa aggagatgca gacctcgtc 900  
gacaaggact gggcatccgt ctttggtctc aagaacggca tgccggaaat caactacaaa 960  
ctctgcgcgc gtaagatgct catgattgac cgcgtcacca gcattgacca caagggcgtt 1020  
gtctacggcc tcggtcagct cgtcggtgaa aagatcctcg agcgcgacca ctggtacttt 1080  
ccctgccact ttgtcaagga tcaggctatg gccggatccc tcgtctccga cggctgcagc 1140  
cagatgtcga agatgtacat gatctggctc ggctccacc tcaccaccgg acctttgac 1200  
ttccgcccgg tcaacggcca cccaacaag gtccgctgcc gcggccaaat cccccgcac 1260  
aagggcaagc tcgtctacgt catggagatc aaggagatgg gcttcgacga ggacaacgac 1320  
ccgtacgcca ttgccgacgt caacatcatt gatgtcgaact tcgaaaaggg ccaggacttt 1380  
agcctcgacc gcacagcga ctacggcaag ggcgacctca acaagaagat cgtcgtcgac 1440  
ttaagggca tcgtctcaa gatgcagaag cgtccacca acaagaaccc ctccaagggt 1500  
cagcccgtct ttgccaacgg cgcgcgcact gtcggccccg aggcctccaa ggcttcctcc 1560  
ggcgccagcg ccagcgccag cgcgcgcgc gccaaagcctg ccttcagcgc cgatgttctt 1620  
gcgcccagc ccgttgccct tcccgagcac atcctcaagg gcgacgcct cgcgcccaag 1680  
gagatgtcct ggcaccccat ggccgcgac ccgggcaacc cgacgcctc ttttgcgccc 1740  
tcggcctaca agccgcgcaa catgccttt acgccttcc ccggcaaccc caacgataac 1800  
gaccacaccc cgggcaagat gccgctcacc tggttcaaca tggccgagtt catggccggc 1860  
aaggtcagca tgtgcctcg ccccgagttc gccaagttcg acgactcgaa caccagccgc 1920  
agccccgctt gggacctcgc tctcgtcacc cgcgcgtgt ctgtgtctga cctcaagcac 1980  
gtcaactacc gcaacatcga cctcgacccc tccaagggtt ccatggtcgg cgagttcgac 2040  
tgccccgcgg acgcctggtt ctacaagggc gcctgcaacg atgccacat gccgtactcg 2100  
atcctcatgg agatcgccct ccagacctcg ggtgtgctca cctcggtgct caaggcgcgc 2160  
ctgaccatgg agaaggacga catcctcttc cgcaacctcg acgccaacgc cgagttcgtg 2220  
cgcgcgcgacc tcgactaccg cggcaagact atccgcaacg tcaccaagtg cactggctac 2280  
agcatgctcg gcgagatggg cgtccaccgc ttcaccttg agctctacgt cgatgatgtg 2340  
ctcttttaca agggctcgac ctcgttcggc tggttcgtgc ccgaggtt ttgcccag 2400  
gccggcctcg acaacggcgc caagtcggag cctggttca ttgagaacaa ggttcggcc 2460  
tcgaggtct cctcctttga cgtgcgcccc aacggcagcg gccgcaccgc catcttcgcc 2520  
aacgccccca gcggcgccca gctcaaccgc cgcacggacc agggccagta cctcgacgcc 2580  
gtcgacattg tctccggcag cggcaagaag agcctcggct acgcccacyg ttccaagacg 2640  
gtcaaccgga acgactggtt cttctcgtgc cacttttgg ttgactcgg catgccgga 2700  
agtctcgggt tcgagtcct gttccagctc gtcgaggcca tcgccccca cgaggatctc 2760  
gctggcaag cacggcattg ccaacccac ctttgtgcac gccccgggc aagatcaagc 2820  
tggaagtacc gcggscagct cacgcccag agcaagaaga tggactcgga ggtccacac 2880  
gtgtccgtgg acgcccacga cggcgttgtc gacctcgtcg ccgacggctt cctctgggc 2940  
gacagcctcc gcgtctactc ggtgagcaac attcgcgtgc gcacgcctc cggtagggcc 3000

cctgcgcgcg cctcctccgc cgcctctgtg ggctcctcgg cttcgtccgt cgagcgcaag 3060  
 cgctcgagcc ccgctgtcgc ctccggcccc gccagacca tcgacctcaa gcagctcaag 3120  
 accgagctcc tcgagctcga tgccccgctc tacctctcgc aggaccgcac cagcggccag 3180  
 ctcaagaagc acaccgacgt ggccctcggc caggccacca tcgtgcagcc ctgcacgctc 3240  
 ggcgacctcg gtgaccgctc cttcatggag acctacggcg tcgtcgcccc gctgtacacg 3300  
 ggcgccaatg ccaagggcat tgccctcggc gacctcgtca tcgccgccgg caagcgcaag 3360  
 atcctcggct cctttggcgc cggcgccctc cccatgcacc acgtgcgcgc cgccctcgag 3420  
 aagatccagg ccgccctgcc tcaggggccc tacgccgtca acctcatcca ctgcctttt 3480  
 gacagcaacc tcgagaaggc caacgtcgat ctcttctcgc agaagggcgt cactgtgggtg 3540  
 gaggcctcgg cattcatgac cctcaccccg caggctcgtc gctaccgcgc cgccggcctc 3600  
 tcgcgcaacg ccgacgggtc ggtaaacatc cgcaaccgca tcatcgggaa ggtctcgcgc 3660  
 accgagctcg ccgagatgtt catccgcccg gcccgggagc acctcctcga gaagctcatc 3720  
 gcctcggggc agatcaccca ggagcaggcc gagctcgcgc gccgcgttcc cgctcgccgac 3780  
 gatatacgtg tcgaggtcga ctccgggggc cacaccgaca accgccccat ccacgtcatc 3840  
 ctccccgtca tcatcaacct ccgcaaccgc ctgcaccgcg agtgccggcta ccccgcgac 3900  
 ctccgcgtcc gcgttggcgc cggcggtggc gtcggctgcc cgcaggccgc cgccgcgcgc 3960  
 ctcaccatgg gcgccgcctt catcgtcacc ggcaactgtca accaggtcgc caagcagtc 4020  
 ggcacctgcg acaacgtgcg caagcagctc tcgcaggcca cctactcgga tatctgcatg 4080  
 gccccggccg ccgacatgtt cgaggagggc gtcaagctcc aggtcctcaa gaagggaacc 4140  
 atgttccctc cgcgcgcaa caagctctac gagctctttt gcaagtacga ctccctcgac 4200  
 tccatgcctc ctgccgagct cgagcgcatc gagaagcgta tcttcaagcg cgactccag 4260  
 gaggtctggg aggagaccaa ggacttttac attaacggtc tcaagaacct ggagaagatc 4320  
 cagcgcgcgc agcacgacct caagctcaag atgtcgtctt gcttccgctg gtaccttgg 4380  
 cttgccagcc gctgggcca catggggcgc ccggaccgcg tcattggacta ccaggctctg 4440  
 tgtggcccg ccattggcgc cttcaacgac ttcatcaagg gcacctacct cgaccccgct 4500  
 gtctccaacg agtacccttg tgtcgtccag atcaacctgc aaatcctccg tgggtgcctgc 4560  
 tacctgcgcc gtctcaacgc cctgcgcaac gaccgcgca ttgacctcga gaccgaggat 4620  
 gctgcctttg tctacgagcc caccaacgcg ctctaagaaa gtgaaccttg tcttaacctg 4680  
 acagcgaatg gcgggagggg gcgggctaaa agatcgtatt acatagtatt tttccctac 4740  
 tctttgtgaa aaaaaaaaa aaaaaaa 4767

&lt;210&gt; 77

&lt;211&gt; 7959

&lt;212&gt; DNA

<213> *Vibrio marinus*

&lt;400&gt; 77

atggctaaaa agaaccacc atcgattaag cagccaagg atgtgttaag tagtgatgat 60  
 caacagttaa attctcgctt gcaagaatgt ccgattgcca tcattggtat ggcatcgggt 120  
 tttgcagatg ctaaaaactt ggatcaattc tgggataaca tcgttgactc tgtggacgct 180  
 attattgatg tgcctagcga tcgctggaac attgacgacc attactcggc tgataaaaaa 240  
 gcagctgaca agacatactg caaacgcggg ggtttcattc cagagcttga ttttgatccg 300  
 atggagtttg gtttaccgcc aaatatcctc gagttaactg acatcgtcga attgttgtca 360  
 ttaattgttg ctctgatgt attaatgat gctggcattg gtagtgatta tgaecatgat 420  
 aaaattggta tcacgctggg tgcgggtggg ggtcagaaac aaatttcgcc attaacgtcg 480  
 cgcctacaag gcccggtatt agaaaaagta ttaaaagcct caggcattga tgaagatgat 540  
 cgcgctatga tcatcgacaa atttaaaaaa gcctacatcg gctgggaaga gaactcattc 600  
 ccaggcatgc taggtaacgt tattgctggg cgtatcgcca atcgttttga ttttgggtgg 660

actaactgtg tgggtgatgc ggcattgcgt ggctcccttg cagctgttaa aatggcgtac 720  
tcagacttac ttgaatatcg ttcagaagtc atgatatcgg gtgggtgatg ttgtgataac 780  
tcgccattca tgtatatgtc attctcgaac acaccagcat ttaccaccaa tgatgatatc 840  
cgtccgtttg atgacgattc aaaaggcatg ctgggtgggtg aagggtattgg catgatggcg 900  
tttaaactgc ttgaagatgc tgaacgtgac ggcgacaaaa ttatttctgt actgaaagg 960  
atcgggtacat cttcagatgg tcgtttcaaa tctattttacg ctccacgccc agatggccaa 1020  
gcaaaagcgc taaaacgtgc ttatgaagat gccgggtttg cccctgaaac atgtgggtcta 1080  
attgaaggcc atgggtacggg taccaaagcg ggtgatgccg cagaatttgc tggttgacc 1140  
aaacactttg gcgccgccag tgatgaaaag caatatatcg ccttaggctc agttaaatcg 1200  
caaattggtc atactaaatc tgcggctggc tctgcgggta tgattaaggc ggcattagcg 1260  
ctgcatcata aaatcttacc tgcaacgac catatcgata aaccaagtga agccttggat 1320  
atcaaaaaca gcccggtata cctaaacagc gaaacgcgtc cttggatgcc acgtgaagat 1380  
gggtattccac gtcgtgcagg tatcagctca tttgggtttg gcggcaccaa cttccatatt 1440  
attttagaag agtatcgccc aggtcacgat agcgcatac gcttaaaactc agtgagccaa 1500  
actgtgttga tctcggcaaa cgaccaacaa ggtattgttg ctgagttaa taactggcgt 1560  
actaaactgg ctgtcgatgc tgatcatcaa ggggtttgtat ttaatgagtt agtgacaacg 1620  
tgcccatata aaaccccatc cgttaaccaa gctcgtttag gttttgttg gcgtaatgca 1680  
aatgaagcga tcgcgatgat tgatacggca ttgaaacaat tcaatgcgaa cgcagataaa 1740  
atgacatggt cagtacctac cgggggtttac tatcgtcaag ccggtattga tgcaacagg 1800  
aaagtgggtg cgctattctc agggcaagg tgcgaatac tgaacatggg tcgtgaatta 1860  
acctgtaact tcccaagcat gatgcacagt gctgcggcga tggataaaga gttcagtgcc 1920  
gctgggttag gccagttatc tgcagttact tccctatcc ctgtttatac ggatgccgag 1980  
cgtaagctac aagaagagca attacgttta acgcaacatg cgcaaccagc gattggtagt 2040  
ttgagtgttg gtctgttcaa aacgtttaag caagcagggt ttaaagctga ttttgctgcc 2100  
gggtcatagt tcggtgagtt aaccgcatta tgggctgccg atgtattgag cgaaagcgat 2160  
tacatgatgt tagcgcgtag tcgtgggtcaa gcaatggctg cgccagagca acaagatttt 2220  
gatgcaggta agatggcgcg tgtgttggt gatccaaagc aagtcgctgt gatcattgat 2280  
accttgatg atgtctctat tgctaacttc aactcgaata accaagttgt tattgctgg 2340  
actacggagc aggttgctgt agcggttaca accttaggtg atgctgggtt caaagttgtg 2400  
ccactgccg tatctgctgc gttccataca ctttagttc gtcacgcgca aaaccattt 2460  
gctaaagcgg ttgatagcgc taaattttaa gcgccaagca ttccagtgt tgctaattgg 2520  
acaggcttgg tgcatcgaag caaacggaat gacattaaga aaaacctgaa aaaccacatg 2580  
ctggaatctg ttcatttcaa tcaagaaatt gacaacatct atgctgatgg tggccgcgta 2640  
tttatcgaat ttggtccaaa gaatgtatta actaaattgg ttgaaaacat tctcactgaa 2700  
aaatctgatg tgactgctat cgcggttaat gctaactcta aacaacctgc ggacgtacaa 2760  
atgcgccaag ctgcgctgca aatggcagtg cttgggtgctg cattagacaa tattgacctg 2820  
tacgacgcg ttaagcgtcc acttgttgcg ccgaaagcat caccaatgtt gatgaagtta 2880  
tctgcagcgt cttatgttag tccgaaaacg aagaaagcgt ttgctgatgc attgactgat 2940  
ggctggactg ttaagcaagc gaaagctgta cctgctgttg tgcacaaacc acaagtgatt 3000  
gaaaagatcg ttgaagttga aaagatagtt gaacgcattg tcgaagtaga gcgtattgtc 3060  
gaagtagaaa aaatcgtcta cgtaaatgct gacgggttcg ttatatcgca aaataatcaa 3120  
gacgttaaca gcgctgtgt tagcaacgtg actaatagct cagtgactca tagcagtgat 3180  
gctgaccttg ttgctctat tgaacgcagt gttgggtcaat ttgttgaca ccaacagcaa 3240  
ttattaaatg tacatgaaca gtttatgcaa ggtccacaag actacgcgaa aacagtgcag 3300  
aacgtacttg ctgcgcagac gagcaatgaa ttaccggaaa gtttagaccg tacattgtct 3360  
atgtataacg agttccaatc agaaacgcta cgtgtacatg aaacgtacct gaacaatcag 3420  
acgagcaaca tgaacaccat gcttactggt gctgaagctg atgtgctagc aacccaata 3480  
actcaggtag tgaatacagc cgttgccact agtcacaagg tagttgctcc agttattgct 3540

aatacagtga cgaatgttgt atctagtgtc agtaataacg cggcgggttc agtgcaaact 3600  
gtggcattag cgcctacgca agaaatcgct ccaacagtcg ctactacgcc agcaccgcga 3660  
ttggttgcta tcgtggctga acctgtgatt gttgcgcatg ttgctacaga agttgcacca 3720  
attacaccat cagttacacc agttgtcgca actcaagcgg ctatcgatgt agcaactatt 3780  
aacaaagtaa tgtagaagt tgttgctgat aaaaccggtt atccaacgga tatgctggaa 3840  
ctgagcatgg acatggaagc tgacttaggt atcgactcaa tcaaactgtg tgagatatta 3900  
ggcgcagtac aggaattgat ccctgactta cctgaactta atcctgaaga tcttgctgag 3960  
ctacgcacgc ttggtgagat tgtcgattac atgaattcaa aagcccaggg ttagctcct 4020  
acaacagtac ctgtaacaag tgcacctgtt tcgcctgcat ctgctggtat tgatttagcc 4080  
cacatccaaa acgtaatgtt agaagtgggt gcagacaaaa ccggttacc aacagacatg 4140  
ctagaactga gcatggatat ggaagctgac ttaggtattg attcaatcaa gcgtgtggaa 4200  
atcttaggtg cagtacagga gatcataact gatttacctg agctaaaccc tgaagatctt 4260  
gctgaattac gcaccctagg tgaaatcggt agttacatgc aaagcaaagc gccagtcgct 4320  
gaaagtgcgc cagtggcgac ggctcctgta gcaacaagct cagcaccgtc tatcgatttg 4380  
aaccacattc aaacagtgat gatggatgta gttgcagata agactggta tccaactgac 4440  
atgctagaac ttggcatgga catggaagct gatttaggta tcgattcaat caaacgtgtg 4500  
gaaatattag gcgcagtgcg ggagatcatc actgatttac ctgagctaaa ccagaagac 4560  
ctcgtgtaat tacgcacgct aggtgaaatc gttagttaca tgcaaagcaa agcgccagtc 4620  
gctgagagtgc gccagtagc gacggcttct gtagcaacaa gctctgcacc gtctatcgat 4680  
ttaaacata tccaaacagt gatgatgaa gtggttgca acaaaaccgg ttatccagta 4740  
gacatgttag aacttgctat ggacatgaa gctgacctag gtatcgattc aatcaagcgt 4800  
gtagaaattt taggtgcggt acaggaatc attactgact tacctgagct taacctgaa 4860  
gatcttgctg aactacgtac attagtgaa atcgtaggt acatgcaaag caaagcgccc 4920  
gtagctgaag cgcctgcagt acctgttgca gtagaaagt cacctactag tgtaacaagc 4980  
tcagcaccgt ctatcgattt agaccacatc caaatgtaa tgatggatgt tgttgctgat 5040  
aagactggtt atcctgcaa tatgcttgaa ttagcaatgg acatggaagc cgaccttgg 5100  
attgattcaa tcaagcgtgt tgaaattcta ggcgcggtac aggagatcat tactgattta 5160  
cctgaactaa acccagaaga cttagctgaa ctacgtacgt tagaagaaat tgtaacctac 5220  
atgcaaagca aggcgagtggt tgttactgta aatgtagtggt ctagccctga aaataatgct 5280  
gtatcagatg catttatgca aagcaatgtg gcgactatca cagcggccgc agaacataag 5340  
gcggaattta aaccggcgcc gagcgcaacc gttgctatct ctcgtctaag ctctatcagt 5400  
aaaataagcc aagattgtaa aggtgctaac gccttaatcg tagctgatgg cactgataat 5460  
gctgtgttac ttgcagacca cctattgcaa actggctgga atgtaactgc attgcaacca 5520  
acttgggtag ctgtaacaac gacgaaagca tttaataagt cagtgaacct ggtgacttta 5580  
aatggcgttg atgaaactga aatcaacaac attattactg ctaacgcaca attggatgca 5640  
gttatctatc tgcacgcaag tagcgaaatt aatgctatcg aataaccaca agcatctaag 5700  
caaggcctga tgtagcctt cttattagcg aaattgagta aagtaactca agccgctaaa 5760  
gtgctggcg cctttatgat tgttactcag cagggtggtt cattaggttt tgatgatatc 5820  
gattctgcta caagtcatga tgtgaaaaca gacctagtag aaagcggtt aaacgggtta 5880  
gttaagacac tgtctcacga gtgggataac gtattctgtc gtgcggttga tattgcttcg 5940  
tcattaacgg ctgaacaagt tgcaagcctt gttagtgtg aactacttga tgctaact 6000  
gtattaacag aagtgggtta tcaacaagct ggtaaaggcc ttgaacgtat cacgttaact 6060  
gggtgtggcta ctgacagcta tgcattaaca gctggcaata acatcgatgc taactcggt 6120  
tttttagtga gtggtggcg aaaaggtgta actgcacatt gtgttgctcg tataagctaaa 6180  
gaatatcagt ctaagttcat cttattggga cgttcaacgt tctcaagtga cgaaccgagc 6240  
tgggcaagtgc gtattactga tgaagcgcg ttaaagaaa cagcgatgca gtctttgatt 6300  
acagcaggtg ataaaccaac acccgtaaag atcgtagc taatcaaacc aatccaagct 6360  
aatcgtgaaa ttgcgcaaac cttgtctgca attaccgctg ctggtggcca agctgaatat 6420

gtttctgcag atgraactaa tgcagcaagc gtacaaatgg cagtcgctcc agctatcgct 6480  
 aagttcgggtg caatcactgg catcattcat ggcgcggggtg tgtagctga ccaattcatt 6540  
 gagcaaaaaa cactgagtga ttttgagtct gtttacagca ctaaaattga cggtttgta 6600  
 tcgctactat cagtcactga agcaagcaac atcaagcaat tggattgtt ctcgtcagcg 6660  
 gctggtttct acggtaaccc cggccagtct gattactcga ttgccaatga gatcttaaat 6720  
 aaaaccgcat accgctttta atcattgcac ccacaagctc aagtattgag ctttaactgg 6780  
 ggctcttggg acgggtggcat ggtaacgcct gagcttaaac gtatgtttga ccaacgtgg 6840  
 gttacatta ttccacttga tgcagggtga cagttattgc tgaatgaact agccgcta 6900  
 gataaccgtt gtccacaaat cctcgtgggt aatgacttat cttaagatgc tagctctgat 6960  
 caaaagtctg atgaaaagag tactgctgta aaaaagccac aagttagtcg tttatcagat 7020  
 gctttagtaa ctaaaagtat caaagcgact aacagtagct ctttatcaaa caagactagt 7080  
 gctttatcag acagtagtgc ttttcagggt aacgaaaacc actttttagc tgaccacatg 7140  
 atcaaaggca atcaggtatt accaacggta tgcgcgattg cttggatgag tgatgcagca 7200  
 aaagcgactt atagtaaccg agactgtgca ttgaagtatg tcggtttcga agactataaa 7260  
 ttgtttaaag gtgtggtttt tgatggcaat gaggcggcgg attaccaa ccaattgtcg 7320  
 cctgtgacaa gggcgctcaga acaggttct gaagtccgta ttgccgcaa gatctttagc 7380  
 ctgaaaagtg acggtaaac tgtgtttcat tatgcagcga caatattgtt agcaactcag 7440  
 ccacttaatg ctgtgaaggt agaacttccg acattgacag aaagtgttga tagcaacaat 7500  
 aaagtaactg atgaagcaca agcgttatac agcaatggca ccttgttcca cggtgaaagt 7560  
 ctgcagggca ttaagcagat attaagttgt gacgacaagg gcctgctatt ggcttgctcag 7620  
 ataaccgatg ttgcaacagc taagcaggga tccttcccgt tagctgacaa caatatcttt 7680  
 gccaatgatt tggtttatca ggctatgttg gtctgggtgc gcaacaatt tggtttaggt 7740  
 agcttacctt cggtgacaac ggcttgact gtgtatcgtg aagtggttgt agatgaagta 7800  
 ttttatctgc aacttaatgt tgttgagcat gatctattgg gttcacgcgg cagtaaaagcc 7860  
 cgttgtgata ttcaattgat tgctgctgat atgcaattac ttgccgaagt gaaatcagcg 7920  
 caagtcagtg tcagtgacat ttgaacgat atgtcatga 7959

&lt;210&gt; 78

&lt;211&gt; 2652

&lt;212&gt; DNA

<213> *Vibrio marinus*

&lt;400&gt; 78

atgacggaat tagctgttat tggtaggat gctaaattta gcggacaaga caatattgac 60  
 cgtgtggaac gcgctttcta tgaagggtct tatgtaggta atgttagccg cgtagtacc 120  
 gaattctaatg ttattagcaa tggcgaagaa caagttatta ctgccatgac agttcttaac 180  
 tctgtcagtc tactagcgca aacgaatcag ttaaatatag ctgatatcgc ggtgttgctg 240  
 attgctgatg taaaaagtgc tgatgatcag cttgtagtcc aaattgcac agcaattgaa 300  
 aaacagtgtg cgagttgtgt tgttattgct gatttaggcc aagcattaaa tcaagtagct 360  
 gatttagtta ataaccaaga ctgtcctgtg gctgtaattg gcatgaataa ctcggttaat 420  
 ttatctcgtc atgatcttga atctgtaact gcaacaatca gctttgatga aaccttcaat 480  
 ggttataaca atgtagctgg gttecgaggt ttacttatcg cttcaactgc gtttgccaat 540  
 gctaagcaat gttatatata cgccaacatt aagggtcttcg ctcaatcggg cgtaaatgct 600  
 caatttaacg ttggaacat tagcgatact gcaagaccg cattgcagca agctagcata 660  
 actgcagagc aggttggttt gttagaagtgc tcagcagtcg ctgattcggc aatcgattg 720  
 tctgaaagcc aaggtttaat gtctgcttat catcatacgc aaactttgca tactgcatta 780  
 agcagtgcc gtagtgtgac tggtagaggc ggggtgtttt cacaggtcgc aggtttattg 840  
 aaatgtgtaa ttggtttaca tcaacgttat attccggcga ttaagattg gcaacaaccg 900

agtgacaatc aaatgtcacg gtggcggaat tcaccattct atatgcctgt agatgctcga 960  
 ccttggttcc cacatgctga tggctctgca cacattgccg cttatagttg tgtgactgct 1020  
 gacagctatt gtcataattct tttaacaaga aacgtcttac aagaacttgt ttgaaagaa 1080  
 acagtcttgc aagataatga cttactgaa agcaagcttc agactcttga acaaaacaat 1140  
 ccagtagctg atctgcgcac taatggttac tttgcatcga gcgagttagc attaatcata 1200  
 gtacaaggta atgacgaagc acaattacgc tgtgaattag aaactattac agggcagtta 1260  
 agtactactg gcataagtac tatcagtatt aaacagatcg cagcagactg ttatgcccg 1320  
 aatgatacta acaaagccta tagcgcagtg cttattgccg agactgctga agagttaagc 1380  
 aaagaaataa ccttggcggt tgctggtatc gctagcgtgt ttaatgaaga tgctaaagaa 1440  
 tggaaaaccc cgaagggcag ttattttacc gcgcagcctg caaataaaca ggctgctaac 1500  
 agcacacaga atgggtgtcac cttcatgtac ccaggtattg gtgctacata tgttggttta 1560  
 gggcggtgac tatttcattct attcccacag atttatcagc ctgtagcggc tttagccgat 1620  
 gacattggcg aaagtctaaa agataactta cttaatccac gcagtattag tcgtcatagc 1680  
 tttaaagaac tcaagcagtt ggatctggac ctgcgcggta acttagccaa tatcgctgaa 1740  
 gccggtgtgg gttttgcttg tgtgtttacc aaggattttg aagaagtctt tgccgttaaa 1800  
 gctgactttg ctacaggtta tagcatgggt gaagtaagca tgtatgcagc actaggctgc 1860  
 tggcagcaac cgggattgat gagtgtcgc cttgcacaat cgaatacctt taatcatcaa 1920  
 ctttgccggc agttaagaac actacgtcag cattggggca tggatgatgt agctaacgg 1980  
 acgttcgagc agatctggga aacctatacc attaaggcaa cgattgaaca ggtcgaaatt 2040  
 gcctctgcag atgaagatcg tgtgtattgc accattatca atacacctga tagcttggtg 2100  
 ttagccggtt atccagaagc ctgtcagcga gtcattaaga atttaggtgt gcgtgcaatg 2160  
 gcaattgaata tggcgaacgc aattcacagc gcgccagctt atgccgaata cgatcatatg 2220  
 gttgagctat accatatgga tgttactcca cgtattaata ccaagatgta ttcaagctca 2280  
 tgttatttac cgattccaca acgcagcaaa gcgatttccc acagtattgc taaatgtttg 2340  
 tgtgatgtgg tggatttccc acgtttggtt aataccttac atgacaaagg tgcgcgggta 2400  
 ttcattgaaa tgggtccagg tcgttcgtta tgtagctggg tagataagat cttagttaat 2460  
 ggcgatggcg ataataaaaa gcaaagccaa catgtatctg ttcctgtgaa tgccaaaggc 2520  
 accagtgatg aacttactta tattcgtgcg attgctaagt taattagtca tggcgtgaat 2580  
 ttgaatttag atagcttggt taacgggtca atcctgggtta aagcaggcca tatagcaaac 2640  
 acgaacaaat ag 2652

&lt;210&gt; 79

&lt;211&gt; 6057

&lt;212&gt; DNA

<213> *Vibrio marinus*

&lt;400&gt; 79

atggatttaa agagagtaat tatggaaaat attgcagtag taggtattgc taatttggtc 60  
 ccggggtcac aagcaccgga tcaattttgg cagcaattgc ttgaacaaca agattgccgc 120  
 agtaaggcga ccgctgttca aatgggcgtt gatcctgcta aatataccgc caacaaagg 180  
 gacacagata aattttactg tgtgcacggc ggttacatca gtgatttcaa ttttgatgct 240  
 tcaggttatc aactcgataa tgattattta gccggtttag atgaccttaa tcaatggggg 300  
 ctttatgtta cgaaacaagc ccttaccgat gcgggttatt ggggcagtac tgactagaa 360  
 aactgtggtg tgatttttag taatttgta ttcccaacta aatcatctaa tcagctgttt 420  
 atgcctttgt atcatcaagt tgttgataat gccttaaagg cggatttaca tcctgatttt 480  
 caattaacgc attacacagc accgaaaaaa acacatgctg acaatgcatt agtagcaggt 540  
 tatccagctg cattgatcgc gcaagcggcg ggtcttggtg gttcacattt tgactggat 600  
 gcggcttggt cttcatcttg ttatagcgtt aagttagcgt gtgattacct gcatacgggt 660



aaagccaaca tgatgcttgc tggcgcggtg tctgcagcag atcctatggt cgtaaataatg 720  
 ggtttctcga tattccaagc ttaccagct aacaatgtac atgccccgtt tgacccaaat 780  
 tcacaaggtc tatttgccgg tgaaggcgcg ggcatgatgg tattgaaacg tcaaagtgat 840  
 gcagtacgtg atggtgatca tatttacgcc attattaaag gcggcgcatc atcgaatgac 900  
 ggtaaaggcg agtttgattt aagcccgaac accaaggggc aagtattagt atatgaacgt 960  
 gcttatgccg atgcagatgt tgaccgagt acagttgact atattgaatg tcatgcaacg 1020  
 ggcacaccta aggggtgacaa tgttggaattg cgttcgatgg aaaccttttt cagtcgcgta 1080  
 aataacaaac cattactggg ctcggttaaa tctaaccttg gtcatttggt aactgccgct 1140  
 ggtagcctg gcatgaccaa agctatgta ggcgtaggta aaggctctat tcctgcaacg 1200  
 attaaactaa agcaaccact gcaatctaaa aacggttact ttactggcga gcaaagcca 1260  
 acgacgactg tgtcttgagg aacaactccg ggtgccaaag cagataaacc gcgtaccgca 1320  
 ggtgtgagcg tatttggttt tgggtggcagc aacgcccatt tgggtattaca acagccaacg 1380  
 caaacactcg agactaattt tagtggtgct aaaccacgtg agcctttggc tattattggt 1440  
 atggacagcc attttggtag tgccagtaat ttagcgcatg tcaaacctt attaaataat 1500  
 aatcaaaata cttcccgta attaccagaa caacgctgga aaggcatgga aagtaacgct 1560  
 aacgtcatgc agtcgttaca attacgcaa gcgcctaaag gcagttacgt tgaacagcta 1620  
 gatattgatt tcttgcggtt taaagtaccg cctaataaaa aagattgctt gatcccgcaa 1680  
 cagttaatga tgatgcaagt ggcagacaat gctgcgaaa acggaggtct agttgaagg 1740  
 cgtaatggtg cgggtattagt agcgatggg atggaactgg aattacatca gtatcggtg 1800  
 cggttaatc taaccacca aattgaagac agcttattac agcaaggat taacctgact 1860  
 gttgagcaac gtgaagaact gaccaatatt gctaaagac gtgttgctc ggctgcacag 1920  
 cttaatcagt atacgagttt cattggtaat attatggcgt cacgtatttc gggttatgg 1980  
 gatttttctg gtcctgctat taccgtatcg gctgaagaaa actctgttta tcgtgtgtt 2040  
 gaattagctg aaaatctatt tcaaacagc gatgttgaag ccgttattat tgctgtgtt 2100  
 gatttgtctg gttcaattga aaacattact ttacgtcagc actacggtc agttaatgaa 2160  
 aaggatctg taagtgaatg tggtcgggtt aatgaaagca gttcagtaac caacaatatt 2220  
 cttgatcagc aacaatggct ggtgggtgaa ggcgcagcgg ctattgtcgt taaacctgca 2280  
 tcgcaagtca ctgctgagca agtttatgcg cgtattgatg cggtaggtt tgcccctggt 2340  
 agcaatgcga aagcaattac gattgcagcg gataaagcat taacacttgc tggtagcgt 2400  
 gctgctgatg tagctagtgt tgaagcacat gcaagtgggt ttagtgccga aaataatgct 2460  
 gaaaaaacg cgttaccgac ttataccca agcgcaagta tcagttcggg gaaagccaat 2520  
 attggtcata cgtttaattg ctcggtatg gcgagtatta taaaacggc gctgctgtta 2580  
 gatcagaata cgagtcaaga tcagaaaagc aaacatattg ctattaacgg tctaggtcgt 2640  
 gataacagct gcgcgcatct tatcttatcg agttcagcgc aagcgcatca agttgcacca 2700  
 gcgcctgtat ctggtatggc caagcaacgc ccacagttag ttaaaacct caaactcgg 2760  
 ggtcagttaa ttagcaacgc gattgttaac agtgcgagtt catctttaca cgctattaaa 2820  
 gcgcagtttg ccggtgaagc cttaacaaa gttaccagc cagtgtatg ggataacctg 2880  
 aagccccag gtattagcgc tcatgcaacc aatgagtatg tggtagctg agctgctaac 2940  
 actcaagctt ctaacattca agcatctcat gttcaagcgt caagtcagc acaagagata 3000  
 gcaccaaacc aagttcaaaa tatgcaagct acagcagcgc ctgtaagtt acccctttct 3060  
 caacatcaac acacagcgca gcccgtagcg gcaccgagcg ttgttgaggt gactgtgaaa 3120  
 cataaagcaa gtaaccaa tcatcagcaa gcgtctacgc ataaagcatt ttagaaagt 3180  
 cgtttagctg cacagaaaaa cctatcgcaa ctgttggaat tgcaaacca gctgtcaatc 3240  
 caaactggta gtgacaatac atctaacaat actgcgtcaa caagcaatac agtgctaaca 3300  
 aatcctgtat cagcaacgac attaacctt gtgtctaag cgctgtagt agcgacaaac 3360  
 ctaaccagta cagaagcaaa agcgcaagca gctgctacac aagctgggtt tcagataaaa 3420  
 ggacctgtt gttacaacta tccaccgctg cagtttaattg aacgttataa taaaccagaa 3480  
 aacgtgattt acgatcaagc tgatttggtt gaattcgctg aaggtgatat tggtaaggta 3540

tttgggtgctg aatacaatat tattgatggc tattcgcgtc gtgtacgtct gccaacctca 3600  
gattacttgt tagtaacacg tgttactgaa cttgatgcc aagtgcatga atacaagaaa 3660  
tcatacatgt gtactgaata tgatgtgcct gttgatgcac cgttcttaat tgatggctag 3720  
atcccttggt ctgttgccgt cgaatcaggc cagtgtgatt tgatgttgat ttcataatc 3780  
gggtattgatt tccaagcgaa aggcgaacgt gtttaccgtt tacttgattg tgaattaact 3840  
ttccttgaag agatggcttt tgggtggcgt actttacgtt acgagatcca cattgattcg 3900  
tatgcacgta acggcgagca attattattc ttcttccatt acgattgtta cgtaggggat 3960  
aagaaggtag ttatcatgag taatgggtgt gctggtttct ttactgacga agaactttct 4020  
gatggtaaag gcgttattca taacgacaaa gacaaagctg agtttagcaa tgctgttaaa 4080  
tcatacattca cgccgttatt acaacataac cgtgggtcaat acgattataa cgacatgatg 4140  
aagttgggta atgggtgatg tgccagttgt tttgggtccgc aatatgatca aggtggccgt 4200  
aatccatcat tgaaattctc gtctgagaag ttcttgatga ttgaacgtat taccaagata 4260  
gacccaaccg gtggtcattg gggactaggc ctgttagaag gtcagaaaaga tttagaccct 4320  
gagcattggg atttcccttg tcactttaaa ggtgatcaag taatggctgg ttcgttgatg 4380  
tcggaagggt gtggccaaat ggcgatgttc ttcattgctg ctcttggtat gcataaccaat 4440  
gtgaacaacg ctggtttcca accactacca ggtgaatcac aaacggtagc ttgtcgtggg 4500  
caagtactgc cacagcgcaa taccttaact taccgtatgg aagttactgc gatgggtatg 4560  
catccacagc cattcatgaa agctaattt gatattttgc ttgacggtaa agtgggtgtt 4620  
gatttcaaaa acttgagcgt gatgatcagc gaacaagatg agcattcaga ttaccctgta 4680  
acactgccga gtaatgtggc gcttaaagcg attactgcac ctgttgcgtc agtagcacca 4740  
gcactctcac ccgctaacag cgcggatcta gacgaacgtg gtgttgaaac gtttaagttt 4800  
cctgaacgtc cgtaaatgcg tgttgagtca gacttgtctg caccgaaaag caaagggtgtg 4860  
acaccgatta agcattttga agcgcctgct gttgctggtc atcatagagt gcctaacca 4920  
gcaccgttta caccttggca tatgtttgag tttgcgacgg gtaatatctc taactgtttc 4980  
ggtcctgatt ttgatgttta tgaaggctgt attccacctc gtacacctg tggcgattta 5040  
caagttgtta ctgaggttgt agaagtgcag ggcgaacgtc ttgatcttaa aaatccatca 5100  
agctgtgtag ctgaatacta tgtaccggaa gacgcttggg actttactaa aaacagccat 5160  
gaaaactgga tgccttattc attaatcatg gaaattgcat tgcaaccaa tggctttatt 5220  
tctggttaca tgggcacgac gcttaaatac cctgaaaaag atctgttctt ccgtaacctt 5280  
gatggtagcg gcacgttatt aaagcagatt gatttacgcg gcaagaccat tgtgaataaa 5340  
tcagtcttg ttagtacggc tattgctggg ggcgcgatta ttcaaagttt cacgtttgat 5400  
atgtctgtag atggcgagct attttatact ggtaaagctg tatttggtta cttagtggt 5460  
gaatcactga ctaaccaact gggcattgat aacggtaaaa cgactaatgc gtggtttgtt 5520  
gataacaata ccccgagc gaattattgat gtgtttgatt taactaatca gtcattggct 5580  
ctgtataaag cgctgtgga taaaccgcat tataaattgg ctggtgggta gatgaacttt 5640  
atcgatacag tgtcagtggt tgaaggcggg ggtaaagcgg gcgtggctta tgtttatggc 5700  
gaacgtacga ttgatgctga tgattggtt ttccgttatc acttccacca agatccggtg 5760  
atgccaggtt cattaggtgt tgaagctatt attgagttga tgcagacctt tgcgtttaa 5820  
aatgatttgg gtggcaagtt tgctaaccga cgtttcattg cgccgatgac gcaagttgat 5880  
tggaataacc gtgggcaaat tacgcccgtg aataaacaga tgtcactgga cgtgcatatc 5940  
actgagatcg tgaatgacgc tgggtgaagt cgaatcgtt gtgatgcga tctgtctaaa 6000  
gatggtctgc gtatttatga agttaaaaac atcgttttta gtattgttga agcgtaa 6057

&lt;210&gt; 80

&lt;211&gt; 1665

&lt;212&gt; DNA

<213> *Vibrio marinus*

&lt;400&gt; 80

atgaatatag taagtaatca ttcggcagct acaaaaaagg aattaagaat gtcgagttaa 60  
 ggttttaaca ataacaacgc aattaactgg gcttggaag tagatccagc gtcagttcat 120  
 acacaagatg cagaaattaa agcagcttta atggatctaa ctaaacctct ctatgtggcg 180  
 aataattcag gcgtaactgg tatagctaata ctagctcag tagcaggtgc gatcagcaat 240  
 aacatcgatg ttgatgtatt ggcgtttgcg caaaagttaa acccagaaga tctgggtgat 300  
 gatgcttaca agaaacagca cggcgtaaa tatgcttata atggcggtgc gatggcaaat 360  
 ggtattgcct cggttgaatt ggttggtgag ttaggttaaag cagggctggt atgttcattt 420  
 ggtgctgcag gtctagtgc ttagtcggtt gaagatgcaa ttcgtcgtat tcaagctgaa 480  
 ttaccaaagtg gcccttatgc ggttaacttg atccatgcac cagcagaaga agcattagag 540  
 cgtggcgagg ttgaacggtt cctaaaactt ggcgtcaaga cggtagaggc ttcagcttac 600  
 cttggtttaa ctgaacacat tgtttggtat cgtgctgctg gtctaactaa aaacgcagat 660  
 ggcagtgtta atatcggtta caaggttatc gctaaagtat cgcgtaccga agttggtcgc 720  
 cgctttatgg aacctgcacc gcaaaaatta ctggataagt tattagaaca aaataagatc 780  
 acccctgaac aagctgcttt agcgttgctt gtacctatgg ctgatgatat tactggggaa 840  
 gcggattctg gtggtcatac agataaccgt cgttttttaa cattattacc gacgattatt 900  
 ggtctgcgtg atgaagtgc agcgaagtat aacttctctc ctgcattacg tgttggtgct 960  
 ggtggtggtg tcggaacgcc tgaagcagca ctgcctgcat ttaacatggg cgcggcttat 1020  
 atcgttctgg gttctgtgaa tcaggcgtgt gttgaagcgg gtgcatctga atataactcg 1080  
 aaactgttat cgacagttga aatggctgat gtgactatgg cacctgctgc agatatgttt 1140  
 gaaatgggtg tgaagctgca agtattaaaa cgcggttcta tgttcgcgat gcgtgcgaag 1200  
 aaactgtatg acttgatgtt ggcttatgac tcgattgaag atatcccagc tgctgaacgt 1260  
 gagaagattg aaaaacaaat ctcccggtgca aacctagacg agatttgagg tggcactatc 1320  
 gctttcttta ctgaacgcga tccagaaatg ctagcccggt caacgagtag tcctaaacgt 1380  
 aaaaaggcac ttatcttccg ttggtatctt ggcctttctt cacgctgggc aaacacaggc 1440  
 gagaagggac gtgaaatgga ttatcagatt tgggcaggcc caagtttagg tgcattcaac 1500  
 agctgggtga aagggtctta cttgaagac tatacccgcc gtggcgctgt agatgttgct 1560  
 ttgcatatgc ttaaagggtg tgcgtattta caacgtgtaa accagttgaa attgcaaggt 1620  
 gttagcttaa gtacagaatt ggcaagttat cgtacgagtg attaa 1665

&lt;210&gt; 81

&lt;211&gt; 2910

&lt;212&gt; DNA

<213> *Shewanella putrefaciens*

&lt;400&gt; 81

atgagtatgt ttttaaattc aaaactttcg cgctcagtca aacttgccat atccgcaggc 60  
 ttaacagcct cgctagctat gcctgttttt gcagaagaaa ctgctgctga agaacaata 120  
 gaaagagtcg cagtgaaccg atcgcaatc gctaaagcag agctaactca accagctcca 180  
 gtcgtcagcc tttcagccga agaactgaca aaatttggtg atcaagattt aggtagcgta 240  
 ctagcagaat tacctgctat tggtgcaacc aacactatta ttggttaata caatagcaac 300  
 tcaagcgtag gtggttagctc agcagacttg cgtcgtctag gtgctaacag aaccttagta 360  
 ttagtcaacg gtaagcgcta cgttgccggc caaccgggct cagctgaggt agatttgta 420  
 actataccaa ctagcatgat ctgcgaggtt gagattgtaa ccggcggtgc ttcagcaatt 480  
 tatgggtcgg acgctgtatc aggtgttata aacgttatcc ttaaagaaga ctttgaaggc 540  
 tttgagttta acgcacgtac tagcgtttct actgaaagt taggcactca agagcactct 600  
 tttgacattt tgggtggtgc aaacgttgca gatggacgtg gtaatgtaac cttctacgca 660  
 ggttatgaac gtacaaaaga agtcatggct accgacattc gccaatcga tgcttgggga 720

acaattaaaa acgaagccga tgggtggtgaa gatgatggta ttccagacag actacgtgta 780  
 ccacgagttt attctgaaat gattaatgct accggtgtta tcaatgcatt tgggtggtgga 840  
 attggtcgct caacctttga cagtaacggc aatcctattg cacaacaaga acgtgatggg 900  
 actaacagct ttgcatttgg ttcatccct aatggctgtg acacatgttt caacactgaa 960  
 gcatacgaaa actatattcc aggggtagaa agaataaacg ttggctcatc attcaacttt 1020  
 gattttaccg ataacattca attttacact gacttcagat atgtaaagtc agatattcag 1080  
 caacaatttc agccttcatt ccgttttggg aacattaata tcaatgttga agataacgcc 1140  
 tttttgaatg acgacttgcg tcagcaaatg ctcgatgcgg gtcaaacc aa tgctagtttt 1200  
 gccaaagttt ttgatgaatt aggaaatcgc tcagcagaaa ataaacgcga acttttccgt 1260  
 tacgtaggtg gcttttaaagg tggctttgat attagcgaaa ccatatttga ttacgacctt 1320  
 tactatgttt atggcgagac taataaccgt cgtaaaacc ttaatgacct aattcctgat 1380  
 aactttgtcg cagctgtcga ctctgttatt gatcctgata ctggcttagc agcgtgtcgc 1440  
 tcacaagtag caagcgctca aggcgatgac tatacagatc ccgcgtctgt aaatggtagc 1500  
 gactgtgttg cttataaccc atttggcatg ggtcaagctt cagcagaagc ccgcgactgg 1560  
 gtttctgctg atgtgactcg tgaagacaaa ataactcaac aagtattgg tggtagcttc 1620  
 ggtaccgatt ctgaagaact atttgagctt caaggtgggt caatcgctat ggttgttgg 1680  
 tttgaatacc gtgaagaaac gtctggttca acaaccgatg aatttactaa agcaggtttc 1740  
 ttgacaagcg ctgcaacgcc agattcttat ggcaataacg acgtgactga gtattttgtt 1800  
 gaggtgaaca tcccagttact aaaagaatta ccttttgcac atgagttgag ctttgacgg 1860  
 gcataccgta atgttgatta ctcacatgcc ggtaagactg aagcatggaa agctggtag 1920  
 ttctactcac cattagagca acttgcatc cgtggtagcg taggtgaagc agtacgagca 1980  
 ccaaacattg cagaagcctt tagtccacgc tctcctggtt ttggccgcgt ttcagatcca 2040  
 tgtgatgcag ataacattaa tgacgatccg gatcgcggtt caaactgtgc agcattgggg 2100  
 atccctccag gattccaagc taatgataac gtcagtgtag ataccttacc tgggtggtaac 2160  
 ccagatctaa aacctgaaac atcaacatcc ttacaggtg gtcttgtttg gacaccaacg 2220  
 tttgtgaca atctatcatt cactgtcgat tattatgata ttcaaattga ggatgctatt 2280  
 ttgtcagtag ccacccagac tgtggctgat aactgtgtt actcaactgg cggacctgac 2340  
 accgacttct gtagtcaagt tgatcgtaat ccaacgacct atgatattga acttgttcgc 2400  
 tctggttacc taaatgccgc ggcattgaat accaaaggta ttgaatttca agctgcatac 2460  
 tcattagatc tagagtcttt caacgcgcct ggtgaactac gcttcaacct attggggaac 2520  
 caattacttg aactagaacg tcttgaattc caaaatcgtc ctgatgagat taatgatgaa 2580  
 aaaggcgaag taggtgatcc agagctgcag ttccgcctag gcacgatta ccgtctagat 2640  
 gatctaagtg ttagctggaa cagcggttat attgatagcg tagtaactta tgatgtctct 2700  
 gaaaatgggtg gctctcctga agatttatat ccaggccaca taggctcaat gacaactcat 2760  
 gacttgagcg ctacatacta catcaatgag aacttcatga ttaacgggtg tgtacgtaac 2820  
 ctatttgacg cacttccacc tggatacact aacgatgcgc tatatgatct agttggtcgc 2880  
 cgtgcattcc taggtattaa ggtaatgatg 2910

&lt;210&gt; 82

&lt;211&gt; 864

&lt;212&gt; DNA

<213> *Shewanella putrefaciens*

&lt;400&gt; 82

atggcaaaaa taaatagtga acacttggat gaagctacta ttacttcgaa taagtgtacg 60  
 caaacagaga ctgaggctcg gcatagaaat gccactacaa cacctgagat gcgccgattc 120  
 atacaagaat cggtatctcag tgtagccaa ctgtctaaaa tattaaatat cagtgaagct 180  
 accgtacgta agtggcgcaa gcgtgactct gtcgaaaact gtcctaatac cccgcacccat 240

ctcaataacca cgctaacccc ttgcaagaa tatgtggttg tgggcctgag ttatcaattg 300  
 aaaatgccat tagacagatt gctcaaagca acccaagagt ttatcaatcc aaacgtgtcg 360  
 cgctcaggtt tagcaagatg ttggaagcgt tatggcggtt cacgggtgag tgatatccaa 420  
 agcccacacg taccaatgag ctactttaat caaattccag tcaactcaag gagcgatgtg 480  
 caaacctaca ccctgcacta tgaacagctg gcaaaaacct tagccttacc tagtaccgat 540  
 ggtgacaatg tgggtgcaagt ggtgtctctc accattccac caaagttaac cgaagaagca 600  
 cccagttcaa ttttgctcgg cattgatcct catagcgact ggatctatct cgacatatac 660  
 caagatggca atacacaagc cacgaataga tatatggctt atgtgctaaa acacggggcca 720  
 ttccatttac gaaagttact cgtgcgtaac tatcacacct ttttacagcg ctttcctgga 780  
 ggcagcgaac atcgccgccc ctctaaagat atgctgaaa caatcaacaa gacgcctgaa 840  
 acacaggcac ccagtggaga ctca 864

<210> 83

<211> 8268

<212> DNA

<213> *Shewanella putrefaciens*

<400> 83

atgagccaga cctctaaacc tacaaactca gcaactgagc aagcacaaga ctcaaacgct 60  
 gactctcggt taaataaacg actaaaagat atgccaatg ctattgttgg catggcgagt 120  
 atttttgcaa actctcgcta ttggaataag ttttgggact taatcagcga aaaaattgat 180  
 gcgattactg aattaccatc aactcactgg cagcctgaag aatattacga cgcagataaa 240  
 accgcagcag acaaaagcta ctgtaaactg ggtggccttt tgccagatgt agacttcaac 300  
 ccaatggagt ttggcctgcc gccaaacatt ttggaactga ccgattcatc gcaactatta 360  
 tcaactcatc ttgctaaaga agtgttggct gatgctaact tacctgagaa ttacgaccgc 420  
 gataaaattg gtatcacctt aggtgtcggc ggtgggtaaa aaattagcca cagcctaaca 480  
 gcgcgtctgc aataccagat attgaagaaa gtattcgcca atagcggcac tagtgacacc 540  
 gacagcgaaa tgcttatcaa gaaattccaa gaccaatatg tacactggga agaaaactcg 600  
 ttcccagggt cacttggtta cgttattgag ggccgtatcg ccaaccgctt cgattttggc 660  
 ggcattgaact gtgtgggtga tgctgcctgt gctggatcac ttgctgctat gcgtatggcg 720  
 ctaacagagc taactgaagg tcgctctgaa atgatgatca ccggtgggtg gtgtactgat 780  
 aactcaccct ctatgtatat gagcttttca aaaacgcccg cctttaccac taacgaaacc 840  
 attcagccat ttgatatcga ctcaaaaggc atgatgattg gtgaaggat ttggcatggtg 900  
 gcgctaaaag gtcttgaaga tgcaagcgc gatggcgacc gcatttactc tgtaattaaa 960  
 ggtgtgggtg catcatctga cggtaagttt aaatcaatct atgcccctcg cccatcaggc 1020  
 caagctaaaag cacttaaccg tgcctatgat gacgcaggtt ttgcgcgcga taccttaggt 1080  
 ctaattgaag ctacaggaac aggtactgca gcaggtgacg cggcagaggt tgccggcctt 1140  
 tgctcagtat ttgctgaagg caacgatacc aagcaacaca ttgcgctagg tttagttaa 1200  
 tcacaaattg gtcatactaa atcaactgca ggtacagcag gtttaattaa agctgctctt 1260  
 gctttgcatc acaagggtact gccgcgcgacc attaacgtta gtcagccaag ccctaaactt 1320  
 gatatacgaac actcaccggt ttatctaaac actgagactc gtccatggtt accacgtgtt 1380  
 gatggtacgc cgcgcgcgcg ggggtattagc tcatttgggt ttggtggcac taacttccat 1440  
 tttgtactag aagagtacaa ccaagaacac agccgtactg atagcgaaaa agctaagtat 1500  
 cgtcaacgcc aagtggcgca aagcttcctt gttagcgcaa gcgataaagc atcgctaatt 1560  
 aacaggttaa acgtactagc agcatctgca agccaagctg agtttatcct caaagatgca 1620  
 gcagcaaaact atggcgtagc tgagcttgat aaaaatgcac cacggatcgg tttagttgca 1680  
 aacacagctg aagagttagc aggcctaatt aagcaagcac ttgccaaact agcagctagc 1740  
 gatgataacg catggcagct acctggtggc actagctacc gcgcgcgtgc agtagaagg 1800

aaagtgtccg cactgtttgc tggccaaggt tcacaatate tcaatatggg ccgtgacctt 1860  
 acttgttatt acccagagat gcgtcagcaa tttgtaactg cagataaagt atttgccgca 1920  
 aatgataaaa cgccgttatc gcaaactctg tatccaaagc ctgtatttaa taaagatgaa 1980  
 ttaaaggctc aagaagccat tttgaccaat accgccaatg cccaaagcgc aattggtgcg 2040  
 atttcaatgg gtcaatacga tttgtttact gcggttggt ttaatgccga catggttgca 2100  
 ggccatagct ttggtgagct aagtgcactg tgtgctgcag gtgttatttc agctgatgac 2160  
 tactacaagc tggcttttgc tcgtggtgag gctatggcaa caaaagcacc ggctaaagac 2220  
 ggcgttgaaag cagatgcagg agcaatgttt gcaatcataa ccaagagtgc tgcagacctt 2280  
 gaaaccgttg aagccaccat cgctaaattt gatggggtga aagtcgctaa ctataacgcg 2340  
 ccaacgcaat cagtaattgc aggcccaaca gcaactaccg ctgatgcggc taaagcgcta 2400  
 actgagcttg gttacaaagc gattaacctg ccagtatcag gtgcattcca cactgaactt 2460  
 gttggtcacg ctcaagcgcc atttgctaaa gcgattgacg cagccaaatt tactaaaaca 2520  
 agccgagcac ttactcaaa tgcaactggc ggactttatg aaagcactgc tgcaaagatt 2580  
 aaagcctcgt ttaagaaaca tatgcttcaa tcagtgcgct ttactagcca gctagaagcc 2640  
 atgtacaacg acggcgcccg tgtatttgtt gaatttggtc caaagaacat cttacaaaaa 2700  
 ttagttcaag gcacgcttgt caacactgaa aatgaagttt gcactatctc tatcaaccct 2760  
 aatcctaaag ttgatagtga tctgcagctt aagcaagcag caatgcagct agcggttact 2820  
 ggtgtgttac tcagtgaat tgaccatac caagccgata ttgccgcacc agcgaaaaag 2880  
 tcgccaatga gcatttcgct taatgctgct aaccatatca gcaaagcaac tcgcgctaag 2940  
 atggccaagt ctttagagac aggtatcgct acctcgcaaa tagaacatgt tattgaagaa 3000  
 aaaatcgctg aagttgagaa actggttgaa gtcgaaaaga tcgtcgaaaa agtggttgaa 3060  
 gtagagaaaag ttgttgaggt tgaagctcct gttaattcag tgcaagccaa tgcaattcaa 3120  
 accggttcag ttgtcgctcc agtaatagag aaccaagtcg tgtctaaaaa cagtaagcca 3180  
 gcagtcacga gcattagtgg tgatgcactc agcaactttt ttgctgcaca gcagcaaac 3240  
 gcacagttgc atcagcagtt cttagctatt ccgcagcaat atggtgagac gttcactacg 3300  
 ctgatgaccg agcaagctaa actggcaagt tctggtgttg caattccaga gagtctgcaa 3360  
 cgctcaatgg agcaattcca ccaactacaa gcgcaaacac tacaagcca caccagttc 3420  
 cttgagatgc aagcgggtag caacattgca gcgttaaacc tactcaatag cagccaagca 3480  
 acttacgctc cagccattca caatgaagcg attcaaagcc aagtggttca aagccaaact 3540  
 gcagtcacgc cagtaatttc aacacaagtt aaccatgtgt cagagcagcc aactcaagct 3600  
 ccagctccaa aagcgcagcc agcacctgtg acaactgcag ttcaaactgc tccggcacia 3660  
 gttgttcgct aagccgcacc agttcaagcc gctattgaac cgattaatac aagtgttgcg 3720  
 actacaacgc cttcagcctt cagcgccgaa acagccctga gcgcaacaaa agtccaagcc 3780  
 actatgcttg aagtgggtgc tgagaaaacc ggttaccxaa ctgaaatgct agagcttgaa 3840  
 atggatatgg aagccgattt aggcacgat tctatcaagc gtgtagaaat tcttggcaca 3900  
 gtacaagatg agctaccggg tctacctgag cttagccctg aagatctagc tgagtgtcga 3960  
 acgctaggcg aaatcggtga ctatatgggc agtaaaactgc cggctgaagg ctctatgaat 4020  
 tctcagctgt ctacaggttc cgcagctgcg actcctgcag cgaatggtct ttctgcggag 4080  
 aaagttcaag cgactatgat gtctgtggtt gccgaaaaga ctggctaccc aactgaaatg 4140  
 ctagagcttg aaatggatat ggaagccgat ttaggcatag attctatcaa gcgcgttgaa 4200  
 attcttggca cagtacaaga tgagctaccg ggtctacctg agcttagccc tgaagatcta 4260  
 gctgagtgtc gtactctagg cgaaatcgtt gactatatga actctaaact cgctgacggc 4320  
 tctaagctgc cggctgaagg ctctatgaat tctcagctgt ctacaagtgc cgcagctgcg 4380  
 actcctgcag cgaatggtct ctctgcggag aaagttcaag cgactatgat gtctgtggtt 4440  
 gccgaaaaga ctggctaccc aactgaaatg ctagaacttg aaatggatat ggaagctgac 4500  
 cttggcatcg attcaatcaa gcgcgttgaa attcttggca cagtacaaga tgagctaccg 4560  
 ggtttacctg agctaaatcc agaagatttg gcagagtgtc gtactcttgg cgaaatcgtg 4620  
 acttatatga actctaaact cgctgacggc tctaagctgc cagctgaagg ctctatgcac 4680

tatcagctgt ctacaagtac cgctgctgctg actcctgtag cgaatggctct ctctgcagaa 4740  
 aaagttcaag cgaccatgat gtctgtagtt gcagataaaa ctggctaccc aactgaaatg 4800  
 cttgaacttg aaatggatat ggaagccgat ttaggtatcg attctatcaa gcgcgttgaa 4860  
 attcttggca cagtacaaga tgagctaccg ggtttacctg agctaaatcc agaagatcta 4920  
 gcagagtgtc gcaccctagg cgaaatcggt gactatatgg gcagtaaaact gccggctgaa 4980  
 ggctctgcta atacaagtgc cgctgctgctt cttaatgtta gtgccgttgc ggcgccctcaa 5040  
 gctgctgcga ctctgtatc gaacggctctc tctgcagaga aagtgc aaag cactatgatg 5100  
 tcagtagttg cagaaaagac cggctaccca actgaaatgc tagaacttgg catggatag 5160  
 gaagccgatt taggtatcga ctcaattaaa cgcgttgaga ttcttggcac agtacaagat 5220  
 gagctaccgg gtctaccaga gcttaatcct gaagatttag ctgagtgcg tacgctgggc 5280  
 gaaatcggtg actatatgaa ctctaagctg gctgacggct ctaagcttcc agctgaaggc 5340  
 tctgctaata caagtgccac tgctgcgact cctgcagtga atggctcttc tgctgacaag 5400  
 gtacaggcga ctatgatgtc tgtagttgct gaaaagaccg gctaccaac tgaaatgcta 5460  
 gaacttggca tggatatgga agcagacctt ggtattgatt ctattaagcg cgttgaaatt 5520  
 cttggcacag tacaagatga gctcccaggt ttacctgagc ttaatcctga agatctcgct 5580  
 gagtgcgcga cgcttggcga aatcgttagc tatatgaact ctcaactggc tgatggctct 5640  
 aaactttcta caagtgcggc tgaaggctct gctgatacaa gtgctgcaa tgctgcaaag 5700  
 ccggcagcaa ttctggcaga accaagtgtt gagcttcctc ctcatagcga ggtagcgcta 5760  
 aaaaagctta atgcggcgaa caagctagaa aattgtttcg ccgcagacgc aagtgttggtg 5820  
 attaacgatg atggctcaca cgcaggcgtt ttagctgaga aacttattaa acaaggccta 5880  
 aaagtagccg ttgtgcgttt accgaaaggt cagcctcaat cgccactttc aagcgatgtt 5940  
 gctagctttg agcttgctc aagccaagaa tctgagcttg aagccagtat cactgcagtt 6000  
 atcgcgcaga ttgaaactca ggttggcgct attggtggct ttattcactt gcaaccagaa 6060  
 gcgaatacag aagagcaaac ggcagtaaac ctatagcgc aaagttttac tcacgttagc 6120  
 aatgcgttct tgtgggcca attattgcaa ccaaagctcg ttgctggagc agatgcgcgt 6180  
 cgctgttttg taacagtaag ccgtatcgac ggtggctttg gttacctaaa tactgacgcc 6240  
 ctaaaagatg ctgagctaaa ccaagcagca ttagctgggt taactaaaac cttaagccat 6300  
 gaatggccac aagtgttctg tcgcgcgcta gatattgcaa cagatgttga tgcaaccat 6360  
 cttgctgatg caatcaccag tgaactattt gatagccaag ctacagctacc tgaagtgggc 6420  
 ttaagcttaa ttgatggcaa agttaaccgc gtaactctag ttgctgctga agctgcagat 6480  
 aaaacagcaa aagcagagct taacagcaca gataaaatct tagtgactgg tggggcaaaa 6540  
 ggggtgacat ttgaatgtgc actggcatta gcatctcgca gccagtctca ctttatctta 6600  
 gctgggcgca gtgaattaca agctttacca agctgggctg agggtaagca aactagcgag 6660  
 ctaaaatcag ctgcaatcgc acatattatt tctactggct aaaagccaac gcctaagcaa 6720  
 gttgaagccg ctgtgtggcc agtgcaaagc agcattgaaa ttaatgccgc cctagccgcc 6780  
 tttacaaaag ttggcgctc agctgaatac gtcagcatgg atgttaccga tagcgccgca 6840  
 atcacagcag cacttaatgg tcgctcaaat gagatcaccg gtcttattca tggcgaggt 6900  
 gtactagccg acaagcatat tcaagacaag actcttgctg aacttgctaa agtttatggc 6960  
 actaaagtca acggcctaaa agcgtgctc gcggcacttg agccaagcaa aattaaatta 7020  
 cttgctatgt tctcatctgc agcaggtttt tacggtaata tcggccaaag cgattacgcg 7080  
 atgtcgaacg atattcttaa caaggcagcg ctgcagttca ccgctcgcaa cccacaagct 7140  
 aaagtcatga gctttaactg gggtccttgg gatggcgga tggtaaaccc agcgcttaaa 7200  
 aagatgttta ccgagcgtgg tgtgtacgtt attccactaa aagcaggtgc agagctatct 7260  
 gccactcagc tattggctga aactggcgtg cagttgctca ttggtacgtc aatgcaaggt 7320  
 ggcagcgaca ctaaagcaac tgagactgct tctgtaaaaa agcttaatgc gggtaggtg 7380  
 ctaagtgcac cgcatccgcg tgctggtgca caaaaaacac cactacaagc tgtcactgca 7440  
 acgctctgt taacccaag tgccatggtc ttcattgaag atcaccgcac tggcggtaac 7500  
 agtgtgttgc caacggtatg cgccatcgac tggatgcgtg aagcggaag cgacatgctt 7560

ggcgctcaag ttaaggtact tgattacaag ctattaaaaag gcattgtatt tgagactgat 7620  
 gagccgcaag agttaacact tgagctaacg ccagacgatt cagacgaagc tacgctacaa 7680  
 gcattaatca gctgtaatgg gcgtccgcaa tacaaggcga cgcttatcag tgataatgcc 7740  
 gatattaagc aacttaacaa gcagtttgat ttaagcgcta aggcgattac cacagcaaaa 7800  
 gagctttata gcaacggcac cttgttccac ggtccgcgtc tacaagggat ccaatctgta 7860  
 gtgcagttcg atgatcaagg ctttaattgct aaagtcgctc tgcctaaggt tgaacttagc 7920  
 gattgtggtg agttcttgcc gcaaaccac atgggtggca gtcaaccttt tgctgaggac 7980  
 ttgctattac aagctatgct gggttgggct cgccttaaaa ctggctcggc aagtttgcca 8040  
 tcaagcattg tgagtttac ctcatacca ccaatggcct ttggtgaaac tggtagcata 8100  
 gagcttgaag tgattaagca caacaaacgc tcacttgaag cgaatgttgc gctatatcgt 8160  
 gacaacggcg agttaagtgc catgtttaag tcagctaaaa tcaccattag caaaagctta 8220  
 aattcagcat tttacctgc tgtcttagca aacgacagtg aggcgaat 8268

<210> 84

<211> 2313

<212> DNA

<213> *Shewanella putrefaciens*

<400> 84

atgccgctgc gcatcgact tatcttactg ccaacaccgc agtttgaagt taactctgtc 60  
 gaccagtcag tattagccag ctatcaaaaca ctgcagcctg agctaaatgc cctgcttaat 120  
 agtgcccgga cacctgaaat gctcagcatc actatctcag atgatagcga tgcaaacagc 180  
 tttgagtcgc agctaaatgc tgcgaccaac gcaattaaca atggctatat cgtcaagctt 240  
 gctacggcaa ctcacgcttt gttaatgctg cctgcattaa aagcggcgca aatgcggatc 300  
 catcctcatg cgcagcttgc cgctatgcag caagctaaat cgacgccaat gagtcaagta 360  
 tctggtgagc taaagcttgg cgctaagcg ctaagcctag ctcagactaa tgcgctgtct 420  
 catgctttta gccaagccaa gcgtaactta actgatgtca gcgtgaatga gtgttttgag 480  
 aacctcaaaa gtgaacagca gttcacagag gtttattcgc ttattcagca acttgctagc 540  
 cgcacccatg tgagaaaaga ggttaatcaa ggtgtggaac ttggccctaa acaagccaaa 600  
 agccactatt ggtttagcga atttcaccaa aaccgtgttg ctgccatcaa ctttattaat 660  
 ggccaacaag caaccagcta tgtgcttact caaggttcag gattgttagc tgcgaaatca 720  
 atgctaaacc agcaaagatt aatgtttatc ttgccgggta acagtcagca acaaataacc 780  
 gcatcaataa ctcagttaat gcagcaatta gagcgtttgc aggttaactga ggttaatgag 840  
 ctttctctag aatgccaaact agagctgctc agcataatgt atgacaactt agtcaacgca 900  
 gacaaactca ctactcgcca tagtaagccc gcttatcagg ctgtgattca agcaagctct 960  
 gttagcgctg caaagcaaga gttaagcgcg cttaacgatg cactcacagc gctgtttgct 1020  
 gagcaaaaaa acgccacatc aacgaataaa ggcttaatcc aatacaaaac accggcgggc 1080  
 agttacttaa ccctaacacc gcttggcagc aacaatgaca acgcccagc gggctcttgc 1140  
 tttgtctatc cgggtgtggg aacggtttac gccgatatgc ttaatgagct gcatcagtac 1200  
 ttccctgcgc ttacgcca aacttgagcgt gaaggcgatt taaaggcgat gctacaagca 1260  
 gaagatatct atcatcttga ccctaaacat gctgcccaaa tgagcttagg tgacttagcc 1320  
 attgctggcg tggggagcag ctacctgtta actcagctgc tcaccgatga gtttaattatt 1380  
 aagcctaatt ttgcattagg ttactcaatg ggtgaagcat caatgtgggc aagcttaggc 1440  
 gtatggcaaa acccgcatgc gctgatcagc aaaacccaaa ccgaccgct atttacttct 1500  
 gctatttccg gcaaatgac cgcggttaga caagcttggc agcttgatga taccgcagcg 1560  
 gaaatccagt ggaatagctt tgtggttaga agtgaagcag cgccgattga agccttgcta 1620  
 aaagattacc cacacgctta cctcgcgatt attcaagggg atacctgcgt aatcgctggc 1680  
 tgtgaaatcc aatgtaaagc gctacttgca gcactgggta aacgcggtat tgcagctaatt 1740



```

cgtgtaacgg cgatgcatac gcagcctgcg atgcaagagc atcaaaatgt gatggatttt 1800
tatctgcaac cgtaaaagc agagcttcct agtgaaataa gctttatcag cgccgctgat 1860
ttaactgcc aagcaaacggg gagtgcgcaa gcacttagca gccaaagtcg tgctcagtct 1920
attgccgaca cttctgcca aaccttgga cttaccgcgc tagtacatca cgcccaacat 1980
caaggcgcta agctgtttgt tgaattggc gcgcatagac aaaactgcac cttgatagac 2040
aagattgtta aacaagatgg tgccagcagt gtacaacatc aaccttggtg cacagtgcct 2100
atgaacgcaa aaggtagcca agatattacc agcgtgatta aagcgcttgg ccaattaatt 2160
agccatcagg tgccattatc ggtgcaacca tttattgatg gactcaagcg cgagctaaca 2220
ctttgccaat tgaccagcca acagctggca gcacatgcaa atgttgacag caagtttgag 2280
tctaaccaag accatttact tcaaggggaa gtc 2313

```

&lt;210&gt; 85

&lt;211&gt; 6012

&lt;212&gt; DNA

<213> *Shewanella putrefaciens*

&lt;400&gt; 85

```

atgtcattac cagacaatgc ttctaaccac cttctgcca accagaaagg cgcatctcag 60
gcaagtaaaa ccagtaagca aagcaaaatc gccattgtcg gtttagccac tctgtatcca 120
gacgctaaaa ccccgcaaga attttggcag aatttgctgg ataaacgcga ctctcgagc 180
accttaacta acgaaaaact cggcgctaac agccaagatt atcaagggtg gcaaggccaa 240
tctgaccgtt tttattgtaa taaaggcggc tacattgaga acttcagctt taatgctgca 300
ggctacaaat tgccggagca aagcttaaat ggcttgagc acagcttcct ttgggcgctc 360
gatactagcc gtaacgcact aattgatgct ggtattgata tcaacggcgc tgatttaagc 420
cgcgaggtg tagtcatggg cgcgctgtcg ttcccaacta cccgctcaaa cgatctgttt 480
ttgccaattt atcacagcgc cgttgaaaaa gccctgcaag ataaactagg cgtaaaggca 540
ttaagctaa gcccaactaa tgctcatacc gctcgcgcgc caaatgagag cagcctaaat 600
gcagccaatg gtgccattgc ccataacagc tcaaaagtgg tggccgatgc acttggcctt 660
ggcggcgcac aactaagcct agatgctgcc tgtgctagtt cggtttactc attaaagctt 720
gcctgcgatt acctaagcac tggcaaagcc gatatcatgc tagcaggcgc agtatctggc 780
gcggatcctt tctttattaa tatgggattc tcaatcttcc acgcctaccc agaccatggt 840
atctcagtac cgtttgatgc cagcagtaaa ggtttggtt ctggcgaaag cgctggcgta 900
ttagtgctta aacgtcttga agatgccgag cgcgacaatg acaaaatcta tgcgggtgtt 960
agcggcgtag gtctatcaaa cgacggtaaa ggccagtttg tattaagccc taatccaaaa 1020
ggtcaggtga aggcctttga acgtgcttat gctgccagtg acattgagcc aaaagacatt 1080
gaagtgattg agtgccacgc aacaggcaca ccgcttggcg ataaaattga gctcacttca 1140
atggaaacct tctttgaaga caagctgcaa ggcaccgatg caccgttaat tggctcagct 1200
aagtctaact taggccacct attaaactgca gcgcatgcgg ggatcatgaa gatgatcttc 1260
gccatgaaag aaggttacct gccgccaagt atcaatatta gtgatgctat cgcttcgccg 1320
aaaaaactct tcggtaaacc aaccctgcct agcatgggtc aaggctggcc agataagcca 1380
tcgaataatc attttggtgt aagaaccgtg cacgcaggcg tatcggtatt tggcttttgt 1440
ggctgtaacg cccatctgtt gcttgagtca tacaacggca aaggaacagt aaaggcagaa 1500
gccactcaag taccgcgtca agctgagccg ctaaaagtgg ttggccttgc ctgcactttt 1560
gggcctctta gcagcattaa tgactcaac aatgctgtga ccaagatgg gaatggcttt 1620
atcgaactgc cgaaaaagcg ctggaaagc cttgaaagc acagtgaact gttagctgaa 1680
tttggttag catctgcgcc aaaagggtgct tatgttgata acttcgagct ggacttttta 1740
cgctttaaac tgccgcaaaa cgaagatgac cgtttgatct cacagcagct aatgctaattg 1800
cgagtaacag acgaagccat tcgtgatgcc aagcttgagc cggggcaaaa agtagctgta 1860

```

ttagtggcaa tggaaactga gcttgaactg catcagttcc gcggccgggt taacttgcatt 1920  
 actcaattag cgcaaagtct tgccgccatg ggcgtgagtt tatcaacgga tgaataccaa 1980  
 gcgcttgaag ccacgcgcac ggacagcgtg cttgatgctg ccaagctcaa tcagtacacc 2040  
 agctttattg gtaatattat ggcgtcacgc gtggcgtcac tatgggactt taatggccca 2100  
 gccttcacta tttcagcagc agagcaatct gtgagccgct gtatcgatgt ggcgcaaaac 2160  
 ctcatcatgg aggataacct agatgcggtg gtgattgcag cggtcgatct ctctggtagc 2220  
 tttgagcaag tcattcttaa aaatgccatt gcacctgtag ccattgagcc aaacctcgaa 2280  
 gcaagcctta atccaacatc agcaagctgg aatgtcgggt aagggtgctgg cgcggtcgtg 2340  
 cttgttaaaa atgaagctac atcgggctgc tcatacggcc aaattgatgc acttggcttt 2400  
 gctaaaactg ccgaaacagc gttggctacc gacaagctac tgagccaaac tgccacagac 2460  
 ttttaataagg ttaagtgat tgaactatg gcagcgctg ctagccaaat tcaattagcg 2520  
 ccaatagtta gctctcaagt gactcacact gctgcagagc agcgtgttg tcaactgctt 2580  
 gctgcagcgg gtatggcaag cctattacac ggcttactta acttaaatat tgtagccaa 2640  
 accaataaag ccaattgcgc gcttatcaac aatatcagtg aaaaccaatt atcacagctg 2700  
 ttgattagcc aaacagcgag cgaacaacaa gcattaaccg cgcgtttaag caatgagctt 2760  
 aaatccgatg ctaaacacca actgggttaag caagtcacct taggtggccg tgatatctac 2820  
 cagcatattg ttgatacacc gcttgcaagc cttgaaagca ttactcagaa attggcgcaa 2880  
 gcgacagcat cgacagtggt caaccaagtt aaacctatta aggcgctggt ctacgtcgaa 2940  
 atggctaact cattcgaaac ggaaagctca gcagagccac aaataacaat tgcagcacia 3000  
 cagactgcaa acattggcgt caccgctcag gcaaccaaac gtgaattagg taccacacca 3060  
 atgacaacaa ataccattgc taatacagca aataatttag acaagactct tgagactgtt 3120  
 gctggcaata ctgttgctag caagggtggc tctggcgaca tagtcaattt tcaacagaac 3180  
 caacaattgg ctcaacaagc tcacctcgcc tttcttgaaa gccgcagtc ggggtatgaag 3240  
 gtggctgatg ctttattgaa gcaacagcta gctcaagtaa caggccaaac tatcgataat 3300  
 caggccctcg atactcaagc cgtcgatact caaacaagcg agaattgtagc gattgccgca 3360  
 gaatcaccag ttcaagttac aacacctgtt caagttacaa cacctgttca aatcagtggt 3420  
 gtggagttaa aaccagatca cgctaattgt ccaccataca cgccgccagt gcctgcatta 3480  
 aagccgtgta tctggaacta tgccgattta gttgagtacg cagaaggcga tatcgccaag 3540  
 gtatttgga gtgattatgc cattatcgac agctactcgc gccgcgtacg tctaccgacc 3600  
 actgactacc tgttggatc gcgcgtgacc aaacttgatg cgaccatcaa tcaatttaag 3660  
 ccatgctcaa tgaccactga gtacgacatc cctgttgatg cgccgtactt agtagacgga 3720  
 caaatccctt gggcggtagc agtagaatca ggccaatgtg acttgatgct tattagctat 3780  
 ctccgtatcg actttgagaa caaaggcgag cggttttacc gactactcga ttgtaccctc 3840  
 accttcttag gcgacttgcc acgtggcgga gataccctac gttacgacat taagatcaat 3900  
 aactatgctc gcaacggcga caccctgctg ttcttcttct cgtatgagt ttttgttggc 3960  
 gacaaagtga tcctcaagat ggatggcggc tgcgtgggt tcttactga tgaagagctt 4020  
 gcggacggta aaggcgtgat tcgcacagaa gaagagatta aagctcgag cctagtgcga 4080  
 aagcaacgct ttaatccgtt actagattgt cctaaaaccc aatttagtta tgggtgatatt 4140  
 cataagctat taactgctga tattgaggtg tgttttggcc caagccacag tggcgctccac 4200  
 cagccgtcac tttgtttcgc atctgaaaaa ttcttgatga ttgaacaagt cagcaagggt 4260  
 gatcgactg gcggtacttg gggacttggc ttaattgagg gtcataagca gcttgaagca 4320  
 gaccactggt acttcccatg tcatttcaag ggcgaccaag tgatggctgg ctcgctaatt 4380  
 gctgaagggt gtggccagtt attgcagttc tatatgctgc accttgggtat gcatacccaa 4440  
 actaaaaatg gtcgtttcca acctcttgaa aacgcctcac agcaagtacg ctgtcgcggt 4500  
 caagtgtgct cacaatcagg cgtgctaact taccgtatgg aagtgactga aatcggtttc 4560  
 agtccacgcc catatgctaa agctaacatc gatattctgc ttaatggcaa agcggtagt 4620  
 gatttccaaa acctaggggt gatgataaaa gaggaagatg agtgactcg ttatccactt 4680  
 ttgactgaat caacaacggc tagcactgca caagtaaacg ctcaacaacg tgcgaaaaag 4740

gtatacaagc cagcatcagc caatgcgcc ttaatggcac aaattcctga tctgactaaa 4800  
 gagccaaaca agggcggtat tccgatttcc catgttgaag caccaattac gccagactac 4860  
 ccgaaccgtg tacctgatac agtgccattc acgccgtatc acatgtttga gtttgctaca 4920  
 ggcaatatcg aaaactgttt cgggccagag ttctcaatct atcgcgcat gatcccacca 4980  
 cgtacaccat gcggtgactt acaagtgacc acacgtgtga ttgaagtaa cggtaagcgt 5040  
 ggcgacttta aaaagccatc atcggtatc gctgaatatg aagtgcctgc agatgcgtgg 5100  
 tatttcgata aaaacagcca cggcgagtg atgccatatt caattttaat ggagatctca 5160  
 ctgcaacctt acggctttat ctgaggttac atgggcacaa ccctaggctt ccctggcctt 5220  
 gagctgttct tccgtaactt agacggtagc ggtgagttac tacgtgaagt agatttacgt 5280  
 ggtaaaaacca tccgtaacga ctacgttta ttatcaacag tgatggccgg cactaacatc 5340  
 atccaaagct ttagcttcga gctaagcact gacggtgagc ctttctatcg cggcactgag 5400  
 gtatttggtt attttaaagg tgacgactt aaagatcagc taggcctaga taacggtaaa 5460  
 gtcactcagc catggcatgt agctaacggc gttgctgcaa gactaagggt gaacctgctt 5520  
 gataagagct gccgtcactt taatgcgcc gctaaccagc cacactatcg tctagccggt 5580  
 ggtcagctga actttatcga cagtgttgaa attgttgata atggcgccac cgaaggttta 5640  
 gggtacttgt atgccgagcg caccattgac ccaagtgtt ggttcttcca gttccacttc 5700  
 caccaagatc cgttatgcc aggtcctta ggtgttgaa caattattga aacctgcaa 5760  
 gcttacgcta ttagtaaaaga cttgggcgca gatttcaaaa atcctaagtt tggtcagatt 5820  
 ttatcgaaca tcaagtggaa gtatcgcggt caaatcaatc cgctgaacaa gcagatgtct 5880  
 atggatgtca gcattacttc aatcaaagat gaagacggt agaaagtcac cacaggtaat 5940  
 gccagcttga gtaaagatgg tctgcgcata tacgaggtct tcgatatagc tatcagcatc 6000  
 gaagaatctg ta 6012

&lt;210&gt; 86

&lt;211&gt; 1629

&lt;212&gt; DNA

<213> *Shewanella putrefaciens*

&lt;400&gt; 86

atgaatccta cagcaactaa cgaaatgctt tctccgtggc catgggctgt gacagagtca 60  
 aatatcagtt ttgacgtgca agtgatggaa caacaactta aagattttag ccgggcatgt 120  
 tacgtggtca atcatgccga ccacggcttt ggtattgcgc aaactgccga tatcgtgact 180  
 gaacaagcgg caaacagcac agatttacct gttagtgtt ttactcctgc attaggtacc 240  
 gaaagcctag gcgacaataa tttccgcgc gttcacggcg ttaaatacgc ttattacgca 300  
 ggcgctatgg caaacggtat ttcattctgaa gagctagtga ttgccctagg tcaagctggc 360  
 attttgtgtg gttcgtttgg agcagccggt cttattccaa gtcgcgttga agcggaatt 420  
 aaccgtattc aagcagcgtt gccaaatggc ctttatatgt ttaaccttat ccatagtcct 480  
 agcgagccag cattagagcg tggcagcgta gagctatttt taaagcataa ggtacgcacc 540  
 gttgaagcat cagctttctt aggtctaaca ccacaaatcg tctattaccg tgcagcagga 600  
 ttgagccgag acgcacaagg taaagttgtg gttggttaaca aggttatcgc taaagtaagt 660  
 cgcaccgaag tggtgaaaa gtttatgatg ccagcgccc caaaaatgct acaaaaaacta 720  
 gttgatgacg gttcaattac cgctgagcaa atggagctgg cgcaacttgt acctatggct 780  
 gacgacatca ctgcagaggc cgattcaggt ggccatactg ataaccgtcc attagtaaca 840  
 ttgctgcaa ccattttagc gctgaaagaa gaaattcaag ctaaatacca atacgacact 900  
 cctattcgtg tcggttggtg tggcggtgtg ggtacgctg atgcagcgt ggcaacgctt 960  
 aacatgggag cggcgatat tgttaccggc tctatcaacc aagcttgtgt tgaagcgggc 1020  
 gcaagtgatc aactcgtaa attacttgcc accactgaaa tggccgatgt gactatggca 1080  
 ccagctgcag atatgttcga gatgggcgta aaactgcagg tggtaagcg cggcacgcta 1140

```
ttcccaatgc gcgctaacaa gctatatgag atctacaccc gttacgattc aatcgaagcg 1200
atcccattag acgagcgtga aaagcttgag aaacaagtat tccgctcaag cctagatgaa 1260
atatgggcag gtacagtggc gcactttaac gagcgcgacc ctaagcaa at cgaacgcgca 1320
gagggtacc ctaagcgtaa aatggcattg attttccgtt ggtacttagg tctttctagt 1380
cgctgggtcaa actcaggcga agtgggtcgt gaaatggatt atcaaatttg ggctggccct 1440
gctctcgggtg catttaacca atgggcaaaa ggcagttact tagataacta tcaagaccga 1500
aatgccgtcg atttggcaaa gcacttaatg tacggcgcgg cttacttaaa tcgtattaac 1560
tcgctaacgg ctcaaggcgt taaagtgcc gcacagttac ttcgctggaa gccaaacca 1620
agaatggc 1629
```

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> <b>C12N 15/52, 15/82, 5/10, 1/21, C12P 7/64, C11C 1/00, C07K 14/405, 14/28, A01H 5/00</b>	<b>A3</b>	<b>(11) International Publication Number:</b> <b>WO 00/42195</b> <b>(43) International Publication Date:</b> 20 July 2000 (20.07.00)
<b>(21) International Application Number:</b> PCT/US00/00956 <b>(22) International Filing Date:</b> 14 January 2000 (14.01.00)  <b>(30) Priority Data:</b> 09/231,899 14 January 1999 (14.01.99) US  <b>(71) Applicant:</b> CALGENE, LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).  <b>(72) Inventors:</b> FACCIOOTTI, Daniel; 2636 Lafayette Drive, Davis, CA 95616 (US). METZ, James, George; 2830 Belhaven Place, Davis, CA 95616 (US). LASSNER, Michael; 721 Falcon Avenue, Davis, CA 95616 (US).  <b>(74) Agent:</b> RAE-VENTER, Barbara; Rae-Venter Law Group, P.C., P.O. Box 60039, Palo Alto, CA 94306 (US).		<b>(81) Designated States:</b> BR, CA, IL, JP, MX, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).  <b>Published</b> <i>With international search report.</i>  <b>(88) Date of publication of the international search report:</b> 28 September 2000 (28.09.00)
<b>(54) Title:</b> SCHIZOCHYTRIUM PKS GENES  <b>(57) Abstract</b>  The present invention relates to compositions and methods for preparing poly-unsaturated long chain fatty acids in plants, plant parts and plant cells, such as leaves, roots, fruits and seeds. Nucleic acid sequences and constructs encoding PKS-like genes required for the poly-unsaturated long chain fatty acid production, including the genes responsible for eicosapentenoic acid production of <i>Shewanella putrefaciens</i> and novel genes associated with the production of docosahexenoic acid in <i>Vibrio marinus</i> are used to generate transgenic plants, plant parts and cells which contain and express one or more transgenes encoding one or more of the PKS-like genes associated with such long chain poly-unsaturated fatty acid production. Expression of the PKS-like genes in the plant system permits the large scale production of poly-unsaturated long chain fatty acids such as eicosapentenoic acid and docosahexenoic acid for modification of the fatty acid profile of plants, plant parts and tissues. Manipulation of the fatty acid profiles allows for the production of commercial quantities of novel plant oils and products.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/00956

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C12N15/52 C12N15/82 C12N5/10 C12N1/21 C12P7/64 C11C1/00 C07K14/405 C07K14/28 A01H5/00		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C12P A01H C07K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 55625 A (CALGENE LLC) 10 December 1998 (1998-12-10)  the whole document	1-3, 5, 7, 17, 18, 20-25, 30-36, 38-40
A	EP 0 823 475 A (NAGASE SEIKAGAKU KOGYO KK ;SUNTORY LTD (JP); JAPAN AS REPRESENTED) 11 February 1998 (1998-02-11) the whole document	4, 6, 8-16, 19, 26-29, 37
A	WO 98 46764 A (THURMOND JENNIFER ;CALGENE LLC (US); ABBOTT LAB (US); KNUTZON DEBO) 22 October 1998 (1998-10-22) see esp. example 20; SEQ ID NO:47-50  --- -/--	4, 6, 8-16, 19, 26-29, 37
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
Date of the actual completion of the international search  14 June 2000		Date of mailing of the international search report  06/07/2000
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer  Kania, T

# INTERNATIONAL SEARCH REPORT

Inter. Application No  
PCT/US 00/00956

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>SOMERVILLE C R: "Future prospects for genetic modification of the composition of edible oils from higher plants; oilseed crop improvement by lipid and fatty acid modification (conference paper)"</p> <p>AMERICAN JOURNAL OF CLINICAL NUTRITION, US, BETHESDA, MD, vol. 58, no. Suppl., 1993, pages 270S-275S, XP002080472</p> <p>see esp. p.274S r. col. 1. par.</p> <p>-----</p>	1-40



# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No  
PCT/US 00/00956

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9855625	A	10-12-1998	EP 1003869 A	31-05-2000
EP 0823475	A	11-02-1998	AU 5346296 A	07-11-1996
			WO 9633263 A	24-10-1996
			JP 2764572 B	11-06-1998
			JP 9000284 A	07-01-1997
			JP 10072590 A	17-03-1998
WO 9846764	A	22-10-1998	US 5972664 A	26-10-1999
			US 5968809 A	19-10-1999
			US 6051754 A	18-04-2000
			AU 7114798 A	11-11-1998
			AU 7114898 A	11-11-1998
			EP 0996732 A	03-05-2000
			EP 1007691 A	14-06-2000
			NO 994924 A	30-11-1999
			NO 994926 A	30-11-1999
			WO 9846765 A	22-10-1998
			AU 6961698 A	11-11-1998
			EP 0975766 A	02-02-2000
			NO 994925 A	30-11-1999
			WO 9846763 A	22-10-1998